

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE VETERINARIA
Departamento de Producción Animal



**ESTRUCTURA GENÉTICA DEL CABALLO DE PURA
RAZA ÁRABE ESPAÑOL Y SU INFLUENCIA EN RAZAS
DERIVADAS: APLICACIÓN DE NUEVAS
METODOLOGÍAS EN EL CÁLCULO DEL TAMAÑO
EFECTIVO.**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Isabel Cervantes Navarro

Bajo la dirección de los doctores
Juan Pablo Gutiérrez García
M^a Mercedes Valera Córdoba

Madrid, 2008

• **ISBN: 978-84-692-1000-0**



UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE VETERINARIA
DEPARTAMENTO DE PRODUCCIÓN ANIMAL

**ESTRUCTURA GENÉTICA DEL CABALLO DE PURA RAZA
ÁRABE ESPAÑOL Y SU INFLUENCIA EN RAZAS
DERIVADAS: APLICACIÓN DE NUEVAS METODOLOGÍAS
EN EL CÁLCULO DEL TAMAÑO EFECTIVO**

POPULATION GENETIC STRUCTURE IN THE SPANISH ARAB HORSE
AND ITS INFLUENCE ON ARAB DERIVED SPANISH HORSE BREEDS:
APPLICATION OF NEW METHODOLOGIES IN THE COMPUTATION OF
THE EFFECTIVE POPULATION SIZE

TESIS DOCTORAL

ISABEL CERVANTES NAVARRO

2008



UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE VETERINARIA
DEPARTAMENTO DE PRODUCCIÓN ANIMAL

**ESTRUCTURA GENÉTICA DEL CABALLO DE PURA RAZA ÁRABE
ESPAÑOL Y SU INFLUENCIA EN RAZAS DERIVADAS:
APLICACIÓN DE NUEVAS METODOLOGÍAS EN EL CÁLCULO
DEL TAMAÑO EFECTIVO**

MEMORIA DE TESIS PRESENTADA POR

ISABEL CERVANTES NAVARRO

Para optar al Grado de Doctor con Mención Europea
por la Universidad Complutense de Madrid

DIRECTORES

Juan Pablo Gutiérrez García

M^a Mercedes Valera Córdoba

Madrid, Octubre de 2008



JUAN PABLO GUTIÉRREZ GARCÍA,
PROFESOR TITULAR DEL DEPARTAMENTO DE PRODUCCIÓN
ANIMAL DE LA FACULTAD DE VETERINARIA DE LA UNIVERSIDAD
COMPLUTENSE DE MADRID

INFORMA:

Que la memoria de Tesis Doctoral titulada: “ESTRUCTURA GENÉTICA DEL CABALLO DE PURA RAZA ÁRABE ESPAÑOL Y SU INFLUENCIA EN RAZAS DERIVADAS: APLICACIÓN DE NUEVAS METODOLOGÍAS EN EL CÁLCULO DEL TAMAÑO EFECTIVO”, realizada por la Licenciada en Veterinaria Dña. Isabel Cervantes Navarro, ha sido realizada bajo su dirección, y cumple las condiciones exigidas para optar al Grado de Doctor por la Universidad Complutense de Madrid.

Para que así conste, firma la presente en Madrid, a uno de Octubre de dos mil ocho.



M^a MERCEDES VALERA CÓRDOBA,
PROFESORA CONTRATADA DOCTORA DEL DEPARTAMENTO DE
CIENCIAS AGROFORESTALES DE LA UNIVERSIDAD DE SEVILLA

INFORMA:

Que la memoria de Tesis Doctoral titulada: “ESTRUCTURA GENÉTICA DEL CABALLO DE PURA RAZA ÁRABE ESPAÑOL Y SU INFLUENCIA EN RAZAS DERIVADAS: APLICACIÓN DE NUEVAS METODOLOGÍAS EN EL CÁLCULO DEL TAMAÑO EFECTIVO”, realizada por la Licenciada en Veterinaria Dña. Isabel Cervantes Navarro, ha sido realizada bajo su dirección, y cumple las condiciones exigidas para optar al Grado de Doctor por la Universidad Complutense de Madrid.

Para que así conste, firma la presente en Madrid, a uno de Octubre de dos mil ocho.

A mi Madre

"Un caballo triunfa cuando se le ha enseñado el camino para ello" (Müsseler)

AGRADECIMIENTOS

Quizás éste debería ser el capítulo más extenso de toda mi Tesis Doctoral, ya que su realización ha sido factible gracias al apoyo de muchas personas. Sin ánimo de querer dejar a nadie atrás, voy a intentar resumir todos mis agradecimientos.

A mi directora y a mi director de tesis. Mercedes, gracias por enseñarme a sacar lo máximo de mí, por tus buenos consejos de maestra y de amiga y por compartir conmigo tus intuiciones que siempre son acertadas. Juan Pablo, gracias por confiar en mí, por el día a día en la Facultad y por todos los conocimientos que compartes conmigo y que espero algún día poder transmitir.

Tan importantes como mis directores han sido otras dos personas. Antonio, gracias por apostar por mí, por tu incondicional apoyo científico y personal y por dar todo por los demás, porque por ello, recibí la parte que me toca cada vez que lo necesito. Y Félix, gracias por la energía positiva que me transmites, por las buenas ideas que compartes conmigo y por empujarme a superarme día a día.

A los cuatro, por ser unos excelentes mentores y abrirme las puertas del conocimiento más allá de nuestra frontera, experiencias que sin duda han contribuido a mi formación. Roswitha, Hans, István e Ino gracias por vuestra atención.

A mi familia, por ser tan grande y maravillosa. A mi padre y a mi madre por apoyarme en todo momento, por ayudarme a crecer en lo personal y en lo laboral y por aguantar mis ausencias. A mi hermano José y a mis hermanas Juana, Leo y Ana, por estar ahí siempre y aportarme los cachitos que completan mi vida, gracias por quererme y demostrármelo.

A mi amiga Cati por ser mi hermana en tierras cordobesas, gracias por tus consejos y por compartir todo conmigo. A mis amigas de siempre, a Rosario y a Yiyi, por estar ahí a pesar de la distancia.

A mi amiga y compañera de batalla, Loli, porque gracias a ella, entré en el mundo de la Genética del caballo. Al resto de compañeros del Grupo MERAGEM: Ester, Pedro, Cristóbal, Alberto, Carmela, Nacho, Francisco, Oliva, Ana, Esther, Manolo y M^a José, gracias por hacerme sentir parte de un gran grupo y por ayudar a hacer más llevaderas las duras tareas.

A todas las personas que contribuyen a que la vida en Madrid sea más fácil, a mis compañeras de piso M^a Isabel y Raquel, a Mónica y a Mar. A mis compañeras de Genética y al resto de personas del Departamento de Producción Animal y de toda la Facultad de Veterinaria de la UCM que me han acogido tan amablemente. Gracias.

Finalmente, un agradecimiento especial a la Asociación Nacional de Criadores de Caballos Árabes, por permitirme trabajar con ellos en el mundo de este magnífico animal.

RESUMEN	1
SUMMARY	7
1. INTRODUCCIÓN	11
2. OBJETIVOS	15
3. REVISIÓN BIBLIOGRÁFICA	17
3.1. EL CABALLO DE PURA RAZA ÁRABE ESPAÑOL	19
3.1.1. Historia del caballo Árabe en España	19
3.1.2. Censos y localización geográfica. La cría en España	21
3.1.3. Descripción morfológica. Prototipo racial	23
3.1.4. Gestión y utilidad de la raza	24
3.1.5. Programa de Mejora	26
3.2. RAZAS DERIVADAS DEL ÁRABE	31
3.2.1. La raza Anglo-Árabe	31
3.2.2. La raza Hispano-Árabe	31
3.2.3. El Caballo de Deporte Español	32
3.3. ESTRUCTURA GENÉTICA DE LA POBLACIÓN	33
3.3.1. Intervalo generacional	33
3.3.2. Grado de profundidad de pedigrí	33
3.3.3. Coeficiente de Consanguinidad (F)	34
3.3.4. Coeficiente de Relación Media (AR)	34
3.3.5. Probabilidad de origen de los genes	35
3.3.5.1. Número efectivo de fundadores (f_e)	36
3.3.5.2. Número efectivo de ancestros (f_a)	36
3.3.5.3. Número efectivo de genomas fundadores (f_g)	36
3.3.6. Estadísticos F de Wright	37
3.4. CÁLCULO DEL CENSO O TAMAÑO EFECTIVO	38
3.4.1. El valor relativo del coeficiente individual de consanguinidad	38
3.4.2. El proceso dispersivo	38
3.4.2.1. El proceso dispersivo desde una óptica de muestreo de alelos	39
3.4.2.2. El proceso dispersivo desde una óptica de consanguinidad	40
3.4.3. Definición de tamaño efectivo	41
3.4.4. La deriva genética y la utilidad del tamaño efectivo	42
3.4.5. Estimación del tamaño efectivo	43
3.4.5.1. Incumplimiento de alguna de las condiciones de la población ideal	43
3.4.5.1.1. Exclusión de la autofecundación	43
3.4.5.1.2. Exclusión del apareamiento fraternal	44
3.4.5.1.3. Número diferente de padres y madres	44
3.4.5.1.4. Número diferente de reproductores en generaciones sucesivas	44

3.4.5.1.5. Distribución no aleatoria del número de hijos por familia	45
3.4.5.1.6. Solapamiento generacional	46
3.4.5.1.7. Mutación, migración y selección	47
3.4.5.2. Estimación del tamaño efectivo en Poblaciones Reales	48
3.4.5.2.1. Estimación del tamaño efectivo en poblaciones con generaciones discretas	48
3.4.5.2.2. Estimación del tamaño efectivo en escenarios de solapamiento generacional	49
3.4.5.3 Estimación del censo efectivo con ausencia de información genealógica	51
3.4.5.3.1. Utilización de la información de campo. Encuestas a ganaderos	51
3.4.5.3.2. Utilización de la información molecular	51
4. ARTÍCULOS	53
4.1. ARTÍCULO 1: POPULATION HISTORY AND GENETIC VARIABILITY IN THE SPANISH ARAB HORSE ASSESSED VIA PEDIGREE ANALYSIS	55
4.2. ARTICULO 2: GENEALOGICAL ANALYSES IN OPEN POPULATIONS: THE CASE OF THREE ARAB-DERIVED SPANISH HORSE BREEDS	73
4.3. ARTÍCULO 3: INDIVIDUAL INCREASE IN INBREEDING ALLOWS ESTIMATING EFFECTIVE SIZES FROM PEDIGREES	93
4.4. ARTICULO 4: APPLICATION OF INDIVIDUAL INCREASE IN INBREEDING TO ESTIMATE REALISED EFFECTIVE SIZES FROM REAL PEDIGREES	111
5. DISCUSIÓN GENERAL	129
5.1. ESTRUCTURA GENÉTICA DEL CABALLO DE PURA RAZA ÁRABE ESPAÑOL Y SU INFLUENCIA EN RAZAS DERIVADAS	131
5.1.1. Estructura Genética del caballo de Pura Raza Árabe Español	131
5.1.1.1. Profundidad de pedigrí	131
5.1.1.2. Relación entre las líneas paternas tradicionales y el análisis relativo a fundadores	132
5.1.1.2.1. Estructuración de la población en cuanto a las líneas paternas estudiada mediante los índices de fijación de Wright	133
5.1.1.3. Las líneas de cría actuales y el análisis relacionado con la consanguinidad	134
5.1.2. Estructura Genética de las razas derivadas del caballo Árabe Español	135
5.1.2.1. Profundidad de pedigrí y consanguinidad	136
5.1.2.2. Influencia de la raza Árabe en sus poblaciones derivadas	136
5.1.2.2. Estimación del censo efectivo en las poblaciones derivadas del caballo Árabe	137
5.1.2.2.1. El tamaño efectivo equivalente sin migración	138

5.2. EL TAMAÑO EFECTIVO REALIZADO. ESTIMACIÓN DEL CENSO EFECTIVO A PARTIR DEL INCREMENTO DE CONSANGUINIDAD INDIVIDUAL	139
5.2.1. El incremento de consanguinidad individual	139
5.2.2. El tamaño efectivo realizado	140
5.2.3. Ensayo del parámetro en poblaciones simuladas	141
5.2.4. Ensayo del parámetro en poblaciones reales	142
5.2.4.1. El tamaño efectivo en el caballo de Pura Raza Árabe Español	143
5.2.5. Limitaciones del tamaño efectivo realizado	144
6. CONCLUSIONES	145
7. REFERENCIAS BIBLIOGRÁFICAS	149
ANEXO: LISTADO DE PUBLICACIONES QUE HA DADO LUGAR LA PRESENTE TESIS DOCTORAL	157

RESÚMENES



RESUMEN

Este trabajo ha sido estructurado en dos bloques. Por un lado se ha analizado la estructura genética del caballo de Pura Raza Árabe (PRÁ) y de sus razas derivadas y, por otro, se ha desarrollado una nueva metodología de estimación del tamaño efectivo y se ha ensayado en poblaciones simuladas y reales.

En primer lugar, se ha estudiado la estructura genética del PRÁ en España para determinar la influencia de los animales fundadores, agrupándolos según su procedencia. Se ha utilizado la información proporcionada por el Libro Genealógico de la raza, desde su fundación en 1847 hasta finales del año 2004, con un total de 18.880 registros. El 15,7% (2.965) de los animales registrados son de origen extranjero. Un 91,9% de los 1.626 animales fundadores fueron individuos identificados como importados de otros países. La suma de las contribuciones genéticas de los fundadores asignados a cada país/región mostraron que el 36,4%, el 18,7%, el 11,9% y el 9,1% de la variabilidad genética se debe a animales procedentes de "Oriente Medio", Polonia, Inglaterra y Egipto, respectivamente. La contribución de individuos fundadores nacidos en España fue considerable (9,9%).

En cuanto a la probabilidad de origen de los genes, se ha encontrado un número efectivo de fundadores de 39,5 para la población de referencia, lo que supone un 4,6% del número total de fundadores encontrados. El número efectivo de ancestros fue de tan solo 13 para los animales nacidos en la última década, mostrando el cuello de botella acaecido durante la Guerra Civil y el uso no equitativo de los animales reproductores. El número de genomas fundadores fue aún inferior con un valor de 6,6.

La estructura de la población, atendiendo a líneas paternas tradicionales indicó que solo un 1% (F_{ST}) de la diferenciación genética de las cuatro subpoblaciones identificadas, se debe a que los individuos se asignen a una línea paterna u otra. Parece ser, que el motivo de la no existencia de una base genética para la diferenciación entre las cuatro subpoblaciones, se debe al uso de determinados individuos no siempre descendientes directos de cada una de las líneas y a la influencia maternal. A todo ello hay que añadir el uso generalizado que se ha realizado de determinados sementales, a lo que han contribuido las paradas públicas de sementales proporcionadas por la Yeguada Militar. Así, la población de PRÁ español presenta una subdivisión sólo parcialmente justificada por la existencia de líneas paternas reconocidas internacionalmente en esta población.

Los coeficientes de consanguinidad mostraron una tendencia ascendente hasta alcanzar 9,8% en los animales nacidos en la última década. El mayor incremento se produce en los años 80, donde observamos que el coeficiente de consanguinidad es superior a la mitad del coeficiente de relación media, lo que indica un apareamiento entre individuos consanguíneos originado por alguna subdivisión en la población, no mostrada en los análisis. Sin embargo, en los últimos diez años los coeficientes de consanguinidad están disminuyendo

probablemente debido a un mayor control en los apareamientos y a la inscripción de animales hijos de sementales extranjeros.

Para analizar la participación del PRÁ en la formación de razas derivadas en España (caballo Anglo-Árabe -A-á-, caballo Hispano-Árabe -H-á- y Caballo de Deporte Español -CDE-) se utilizaron los Libros Genealógicos de estas 4 razas junto a los de las otras razas parentales (caballo de Pura Raza Español -PRE- y el caballo de Pura Sangre Inglés -PSI-) contando con un total de 211.754 individuos. Se compararon los resultados obtenidos con las bases de datos exclusivas de las razas derivadas (en cuyo Libro Genealógico no están incluidas las genealogías de animales puros que actúan como reproductores) con los obtenidos tras reconstruir las genealogías de animales de las razas parentales.

El análisis de la influencia del PRÁ en la formación de estas razas se realizó de dos formas. Primero, sumando las contribuciones de los animales fundadores por raza obteniéndose que para la raza A-á e H-á, la contribución de la raza Árabe fue siempre inferior a la de la otra raza parental (PSI o PRE) y tendió a ser mayor por la vía materna. En el caso del CDE, la contribución de fundadores PRÁ fue inferior al 10%. Este método reflejó las preferencias de los criadores a la hora de determinar los cruzamientos. El segundo método se realizó utilizando la matriz de coascendencias, que englobaba las bases de datos de las 6 razas. Se obtuvo que por término medio un individuo de la raza A-á comparte el 4,6% de sus genes con un caballo de la población actual de PRÁ, mientras que un individuo H-á comparte el 6,4% con otro animal de la población actual de PRÁ. El porcentaje de genes compartidos con la otra raza parental (PSI y PRE) es en ambos casos inferior. Se concluye que, aunque los criadores tienden a preferir la otra raza parental para la creación de animales de razas derivadas, los sementales utilizados en PRÁ y en las razas derivadas parecen ser los mismos, no dándose este hecho con la otra raza parental.

Se utilizó la teoría del modelo isla en las bases de datos exclusivas de las poblaciones derivadas para desarrollar el concepto de tamaño efectivo equivalente sin migración. Los censos efectivos equivalentes estimados de este modo fueron 39,1, 56,3 y 114,1 para A-á, H-á y CDE respectivamente, que pueden ser comparables a los obtenidos en poblaciones puras. Esto permitió establecer comparaciones entre poblaciones en términos de censo y política de apareamientos, resultando el A-á comparable al PRÁ y el H-á al PRE.

Se desarrolló también un método de estimación del tamaño efectivo útil en situaciones en las que los métodos convencionales no se ajustan de forma adecuada. La nueva metodología se basó en el uso de los incrementos de consanguinidad individuales (ΔF_i). A partir del inverso del doble del valor medio de los incrementos de consanguinidad individuales de la población de referencia que se desea estudiar, se define el parámetro que hemos llamado Tamaño o Censo Efectivo Realizado. El parámetro se acompaña de una medida de error. La metodología fue aplicada en ejemplos simulados y en poblaciones reales comprobando la validez y utilidad de los resultados.

En cuanto a los ejemplos simulados, la metodología fue puesta a prueba en circunstancias habituales de las poblaciones de animales domésticos, como son la subdivisión, el solapamiento o cuando el número de descendientes por individuo no sigue la esperada distribución de Poisson. Se produjo una ligera sobreestimación del verdadero valor del parámetro que se va reduciendo a medida que avanzan las generaciones probablemente debida al retraso que se produce en poblaciones dioicas al excluirse la autofecundación. La variabilidad fue elevada en situaciones en las que su propio valor era grande, presentando un efecto de escala. En el caso de la subdivisión, el parámetro se ajusta de acuerdo a la historia completa de la población. Si se desea conocer el efecto de un cambio de política de apareamientos, conviene utilizar un pedigrí truncado en el momento en el que este cambio fue instaurado.

En cuanto al ensayo en poblaciones reales, podemos destacar su estabilidad y su utilidad cuando otras metodologías no proporcionan valores coherentes. En el caso de poblaciones bajo el efecto de la migración su comportamiento fue también adecuado obteniéndose también un elevado error estándar, debido a que en poblaciones con entrada constante de nuevos individuos procedentes de otros Libros Genealógicos, el número de generaciones conocidas es muy reducido. El tamaño efectivo realizado en el caballo de PRá fue de 39,1.

La metodología ha demostrado que puede ser aplicada en cualquier escenario en el que se disponga de datos genealógicos, al contrario que metodologías comúnmente utilizadas que no conducen a resultados satisfactorios en determinadas situaciones.

SUMMARY

This work has been structured in two parts. In the first part, the genetic structure of the Spanish Arab horse and its derived-breeds was analysed. In the second part, a new methodology was developed to estimate the effective population size; it was tested in simulated and real populations.

The genetic structure of the Spanish Arab horse was first analysed to determine the influence of the founders, grouping them according to their country origin. The information provided by the Spanish Arab Studbook was used, from its foundation in 1847 to 2004, with a total of 18,880 individuals. The 15.7% (2,965) of the total registered animals are of foreign origin. A 91.9% of the 1,626 founders was identified as foreign animals. The sum of the genetic contributions of the founders assigned to each country/area showed that 36.4%, 18.7%, 11.9% and 9.1% of the genetic variability is explained by animals coming from the "Middle East", Poland, England and Egypt, respectively. The contribution of founders born in Spain was considerable at 9.9%.

Regarding the probability of gene origin, the effective number of founders was 39.5 for the reference population, which supposes a 4.6% of the total number of founders. The effective number of ancestors was 13 for the animals born in the last decade, showing that a bottleneck happened during the Civil War and the unbalanced use of the reproductive animals. The number of founder genome equivalents was still lower with a value of 6.6.

The structure of the population, by traditional paternal lines indicated that only a 1% (F_{ST}) of the genetic differentiation between the four identified subpopulations, is due to individuals being assigned to one paternal line or another. It seems, that the reason for the nonexistence of a genetic base for the differentiation between the four subpopulations are the use of no direct descendants of each line, the maternal influence and the generalized use of specific stallions, to which contributed the "Paradas Militares" provided by the Military Stud. Thus, the population of Spanish Arab Horse presents a subdivision only partially justified by the existence of paternal lines recognized internationally.

The inbreeding coefficients showed an ascending trend until reaching 9.8% in the animals born in the last decade. The greater increase took place in the 1980's, where we observed that the inbreeding coefficient is higher than half average relatedness coefficient, which indicates a mating between related individuals, originated by some subdivision in the population, not shown in the analyses. Nevertheless, in the last decade the inbreeding coefficient is decreasing, probably due to a higher control in the breeding policy and to the use of foreign stallions.

In order to quantify the influence of the Spanish Arab Horse in the formation of its derived-breeds in Spain (Anglo-Arab horse, Hispano-Arab horse and Spanish Sport Horse) the Studbooks of these 4 breeds and those of the other parental breeds (Spanish Purebred horse and Spanish Thoroughbred) were collected making a total of 211,754 individuals. A comparison was made of the results

obtained with the breed exclusive datasets of the derived-breeds (whose Studbooks do not include the genealogies from pure animals that act as stallions/mares), and those obtained after completing the animal genealogies (completed datasets) from parental breeds.

The analysis of the influence of Spanish Arab horse in the formation of these breeds was made in two ways. First, adding the contributions of the founders by breed, established that for Anglo-Arab and Hispano-Arab the contribution of the Arab horse was always lower than the other parental breeds (Spanish purebred or Thoroughbred) and tended to be higher on the maternal side. In the case of the Spanish Sport horse, the contribution of Arab founders was lower than 10%. This method showed the preferences of the breeders at the time of breeding the derived individuals. In the second method, we used the coancestry matrix that included the datasets of the six breeds. From the results it was established that an Anglo-Arab individual shares 4.6% of its genes with a horse of the present population of the Arab horse, whereas a Hispano-Arab individual shares 6.4% of the genes. The percentage of genes shared with the other parental breeds (Spanish purebred or Thoroughbred) was lower in both cases. We can conclude that, although the breeders tend to prefer the other parental breed to create animals for derived breeds, the stallions used in Spanish Arab Horse and in the derived breeds seem to be the same. This fact does not occur with the other parental breeds.

The island model theory was applied in the breed exclusive datasets of the derived populations to develop the concept of equivalent effective size without migration. The equivalent effective sizes were 39.1, 56.3 and 114.1 for the Anglo-Arab, Hispano-Arab and Spanish Sports horse, respectively. These values are comparable with those obtained in pure populations. This enabled us to make comparisons between populations in terms of census and breeding policy: Anglo-Arab to the Arab and the Hispano-Arab to the Spanish Purebred horse.

We developed in this study a method to obtain reliable estimates of effective population size, useful under situations in which the conventional methods do not fit properly. The new methodology is based on the use of the individual increases in inbreeding (ΔF_i). From the inverse of the double of the average value of the individual increase in inbreeding from the reference population, we obtained the parameter that we called Realised Effective Population Size. A uncertainty measure was also developed. The methodology was applied in simulated examples and real populations verifying the validity and utility of the results.

The methodology is demonstrated by applying it to four simulated examples embracing a wide range of typical theoretical scenarios: subdivision, overlapping or when the number of descendants does not follow the Poisson distribution. A slight overestimation of the true value of the parameter occurred at the beginning but it was reduced as the generations advanced, probably due to the delay that takes place in populations without self-fertilization. The variability was high in situations in which the parameter value was large, giving a scale effect. In the case of the subdivision, the parameter is adjusted according to the complete history of the population. If the aim for estimating the effective population size is to ascertain the

effect of changes in the mating policy, it would be recommendable to truncate the pedigree, thus considering as founders those animals which were used to establish the new mating policy.

Regarding real population examples, we can emphasize the stability and utility of the method when other methodologies do not provide coherent values. In the case of populations under the effect of migration the behaviour was also adapted resulting in a high standard error, because in populations with a constant entrance of new individuals coming from other Studbooks, the number of equivalent generations is low. The realised effective size for Spanish Arab horse was 39.1.

Additionally this methodology can be applied to any scenario with genealogical data, where the classical methods do not lead to satisfactory results in certain situations.

INTRODUCCIÓN Y OBJETIVOS



1. INTRODUCCIÓN

El caballo de Pura Raza Árabe (PRÁ) constituye una de las razas más importantes de mundo, participando en la formación de muchas otras, ya haya sido en su origen (Pura Sangre Inglés, Bowling y Ruvinsky, 2000 o Lipizzano, Zechner *et al.*, 2002) o por su continuo uso en el cruzamiento para la formación de razas derivadas, de las cuales en España destacan el caballo Anglo-Árabe, el caballo Hispano-Árabe y el caballo de Deporte Español.

Su distribución mundial conlleva que existan diversas “poblaciones” de caballos Árabes, cada una de ellas con Libro Genealógico propio. Según la WAHO (World Arabian Horse Organisation), organización internacional bajo la que se encuentran los Libros Genealógicos reconocidos de esta Raza, el Caballo Árabe Puro es “aquél que aparece en cualquier Studbook o Registro reconocido para la raza”. De este modo, todos los Libros Genealógicos de la raza permanecen abiertos a la inclusión de individuos Pura Raza Árabe inscritos en Libros Genealógicos de otros países, manteniéndose un flujo constante entre ellos.

El PRÁ es una raza pura, pero a su vez con diferentes Libros Genealógicos, todos ellos englobados y supervisados por la WAHO y debido al intercambio producido entre los animales inscritos en los diferentes Libros. Este hecho le dota de unas características diferenciales a otras razas puras caballares que podemos encontrar. Además, su participación en la formación de razas derivadas le permite ser el hilo de conexión entre todas ellas.

Entre las poblaciones mundiales de caballos Árabes con Libro Genealógico en funcionamiento, el PRÁ español constituye una de las más importantes. Su extraordinaria resistencia y su morfología hacen que este animal sea muy deseable a los ojos de los aficionados a las competiciones deportivas y de los criadores de ganado equino tanto para su cría en pureza, como para sus razas derivadas. Por todo ello, su censo ha aumentado continuamente desde la fundación de su libro genealógico en 1847 y de forma manifiesta en los últimos años (Valera *et al.*, 2007).

Desde 2002, el marco del sector equino español se ha visto incentivado por el desarrollo de un Plan Nacional de Ordenación y Fomento del sector equino, en el que se ha trabajado desde el sector ministerial legislando los Libros Genealógicos, las características zootécnicas de las razas, los controles de rendimientos, etc., y desde el punto de vista científico, donde se trabaja en el diseño de Planes de Mejora y Esquemas de Selección de las distintas razas, profundizando en las investigaciones en el campo de la genética, la fisiología, la locomoción y la morfoestructura equina.

Para el establecimiento de programas que gestionen de manera óptima la variación genética de la raza, así como para el desarrollo de sus esquemas de selección genética, es necesario como paso previo el análisis de la estructura genética de las poblaciones animales, así como el conocimiento de la variabilidad genética y el modo en que se produce el flujo de genes.

Para realizar el estudio de la estructura genética de una población se hace necesario el uso de parámetros que nos permitan conocer cómo se ha llevado a cabo el flujo de genes en la misma, proporcionando información sobre el número de genes que la ha originado y estimando en qué medida participan en la población actual. De entre todos los parámetros útiles en la descripción de la estructura genética de poblaciones animales destaca el tamaño o censo efectivo. Debido a su relación directa con la tasa de consanguinidad, (Falconer y Mackay, 1996) se considera como un parámetro imprescindible en el campo de la conservación y genética de poblaciones (FAO, 1998, Duchev *et al.*, 2006), ya que proporciona información sobre el grado de variabilidad genética de una población y permite tomar decisiones en relación a su gestión genética.

Esta Tesis Doctoral se estructura en dos partes conectadas, pero diferenciadas. En la primera parte se realiza un estudio poblacional del caballo de Pura Raza Árabe y su influencia en la formación de sus razas derivadas: caballo Anglo-Árabe español, caballo Hispano-Árabe y caballo de Deporte Español, mediante las herramientas disponibles en la metodología utilizada en la descripción de la estructura genética de poblaciones. En esta primera parte, se ponen de manifiesto las dificultades que podemos encontrar en la utilización de las metodologías clásicas, para la determinación del tamaño efectivo cuando se trata de poblaciones sometidas a migración y cuando la consanguinidad evoluciona de forma irregular ante cambios en la política de apareamientos. En consecuencia, en la segunda parte de la Tesis se propone una metodología alternativa útil en este tipo de escenarios, pero también perfectamente aplicable en cualquier otro. Se desarrolla en esta segunda sección una nueva metodología de estimación de este parámetro genealógico basado en incrementos de consanguinidad individual y su aplicación en diversas poblaciones, entre ellas, y como es lógico en el Caballo de Pura Raza Árabe Español.

2. OBJETIVOS

El objetivo general de esta Tesis Doctoral ha sido el estudio poblacional del caballo de Pura Raza Árabe Español y su contribución genética a sus razas derivadas en España (caballo Anglo-Árabe, caballo Hispano-Árabe y Caballo de Deporte Español). Dadas las características de las poblaciones estudiadas se hizo necesario además, el desarrollo y puesta a punto de una metodología para la estimación del tamaño efectivo poblacional. El objetivo general se puede desglosar en los siguientes objetivos específicos:

1. Análisis de la estructura genética del caballo de Pura Raza Árabe Español mediante análisis genealógico.
2. Estudio de la influencia de la raza Árabe en la formación de sus principales razas derivadas en España: el caballo Anglo-Árabe, el caballo Hispano-Árabe y el Caballo de Deporte Español.
3. Desarrollo de una metodología apropiada para estudiar la estructura genética de poblaciones sometidas a migración.
4. Desarrollo de una nueva metodología para la estimación del tamaño efectivo en poblaciones reales de animales con registro genealógico.
5. Ensayo de la metodología desarrollada en el objetivo previo, en poblaciones modelo, tanto simuladas como reales.

REVISIÓN BIBLIOGRÁFICA



3. REVISIÓN BIBLIOGRÁFICA

3.1. EL CABALLO DE PURA RAZA ÁRABE ESPAÑOL

3.1.1. Historia del caballo Árabe en España

La cría del caballo Árabe se inicia durante la invasión musulmana (siglos VIII a XV), de forma que España es el primer país europeo donde entra el caballo de Pura Raza Árabe (PRÁ).

Desde el principio es el ejército el encargado de su cría, adquiriendo caballos del desierto, comprados o arrebatados como botín de guerra en los pueblos nómadas durante los siglos XV-XVIII contribuyendo a la mejora del caballo Árabe Español.

El primer depósito permanente de sementales se crea en 1834 (siglo XIX) en Córdoba. Más tarde, en 1847, el Libro Genealógico de la raza es creado por el Departamento de Cría Caballar del Ejército Español, registrando junto con individuos PRÁ otras razas como el caballo de Pura Raza Español (PRE), Pura Sangre Inglés (PSI) y caballo Anglo-Árabe (A-á) (Maxwell, 1995). En ese mismo año la Yeguada Militar se traslada a Jerez de la Frontera (Cádiz).

Los primeros PRÁ registrados fueron importados procedentes del desierto en nombre de la Reina Isabel II. Las siguientes importaciones se realizaron por cuenta del Gobierno Español, por el Duque de San Carlos y por el General Marqués. Dicha importación sirvió sobre todo, para mejorar las razas existentes y ninguna de sus líneas ha sobrevivido hasta nuestros días (Maxwell, 1995).

En la sección del PRÁ del Volumen I del Libro de Registro Racial que se publica en 1885 aparecen 36 sementales (30 procedentes del desierto, 4 de Francia, 1 de la India y 1 de España) y 12 yeguas procedentes del desierto. Los siguientes volúmenes (años 1886-1903) incluyen animales importados, sobre todo de Francia, pero ninguno fue usado para la cría en pureza.

En 1893, la Yeguada Militar, que depende del entonces denominado Ministerio de la Guerra, se traslada de nuevo a Córdoba. En este momento se impone un control drástico para la matriculación en el Libro Genealógico para preservar la calidad de los ejemplares inscritos en el mismo. A los caballos importados se les exigen documentos de orígenes y pruebas de linaje para poder ser inscritos. Debido a las dificultades para hallar ejemplares de calidad, el ejército decide organizar en España una cría del caballo Árabe.

A principios de siglo XX se publica el volumen XI del Libro de Registro del PRÁ donde se recogen las importaciones realizadas por Yeguada Militar, principalmente del desierto y de Polonia, dando lugar a la mayoría de los ancestros del caballo de Pura Raza Árabe en España.

El primer caballo Árabe nacido en España ("Pure Spanish", como se denomina en el extranjero los ejemplares españoles), *Nakib*, fue registrado en el volumen XII (Maxwell, 1995). La Yeguada Militar siguió haciendo importaciones de Francia,

pero las más importantes se realizaron entre 1908 y 1912, dando lugar a las líneas que persisten hoy día.

La cría privada comenzó en 1910 con la ganadería del Marqués de Casa Domecq seguida de José María y Juan del Cid. En 1917 Ybarra comienza la compra de yeguas y en 1920 funda "La Cascajera" en Sevilla. Más tarde, su hijo Luis, tomará el relevo. El Duque de Veragua comienza las importaciones y la cría en 1920, al igual que los Hermanos Guerrero realizando importaciones en 1923 de Francia y de Egipto.

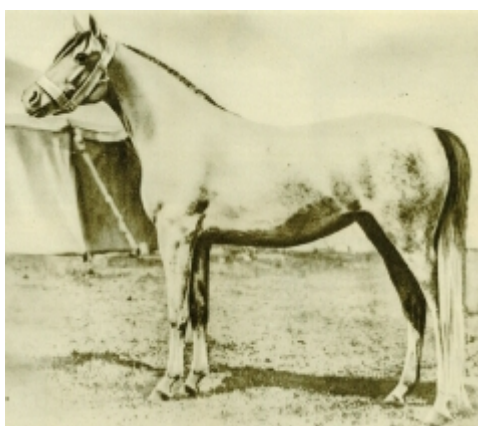
De entre los primeros criadores privados destaca especialmente el Duque de Veragua (Steen, 2006). Su aportación a la cría del PRá Español ha sido muy importante tanto en calidad como cantidad de ejemplares. Entre 1926 y 1934 importó 13 yeguas y 2 sementales de Inglaterra, realizó importaciones de Argentina, y en 1935, tras la muerte del Marqués de Domecq, adquirió las yeguas de su ganadería. Su yeguada estaba repartida en dos lugares diferentes: Madrid, donde se situaban los sementales y Valjuanete, finca situada entre Toledo y Aranjuez, donde se criaban las yeguas y potros.

La Guerra Civil Española (1936-1939) afectó en gran medida al sector equino en España. Muchas ganaderías vieron disminuidos sus efectivos, se destruyeron documentos y se requisaron caballos para el frente.

En 1939, al final de la guerra, los caballos supervivientes fueron devueltos a sus criadores exceptuando los del Duque de Veragua. Los animales situados en Valjuanete fueron rescatados por el ejército y trasladados al depósito de Moratella en Córdoba.

Otros muchos ejemplares se vendieron a particulares. Muchas yeguas fueron llevadas a la Yeguada Militar y las potras que no fueron identificadas debido a la destrucción de documentos de la ganadería, fueron renombradas usando el prefijo "Vera". Otro grupo de yeguas fue a parar a Francisco Montero de Espinosa; éstas no recibieron el prefijo "Vera". A ninguno de los potros se les puso el prefijo. Algunas potritas descendientes de *Caftan* 1925 fueron vendidas a la Marquesa de Avella que era la nieta del duque de Veragua y que continuó la cría en Valjuanete (Steen, 2006).

Después de la Guerra Civil, el caballo Árabe español había perdido gran parte de su Libro Genealógico dejando de suscitar interés en el mercado internacional. El aislamiento de España tras la Guerra produjo también un cese de las importaciones.



Reyna, yegua PRá, importada de Inglaterra por el Duque de Veragua. Reproducción de 1950 (Yeguada Militar de Córdoba).

La cría de la raza no se recuperó hasta 1970, cuando la World Arabian Horse Organization (WAHO) aceptó al caballo Árabe español dando el visto bueno a la supuesta genealogía del Duque de Veragua (Llamas *et al.*, 1992).

De nuevo la Yeguada Militar tuvo un papel importante y la cría se expandió en los años 80, pero la aparición de la peste equina en España en 1987 impidió exportaciones hasta 1993. Esto no afectó a las importaciones que continuaron trayendo animales procedentes de Egipto, principalmente líneas maternas egipcias, de Polonia y de Inglaterra (Llamas *et al.*, 1992).

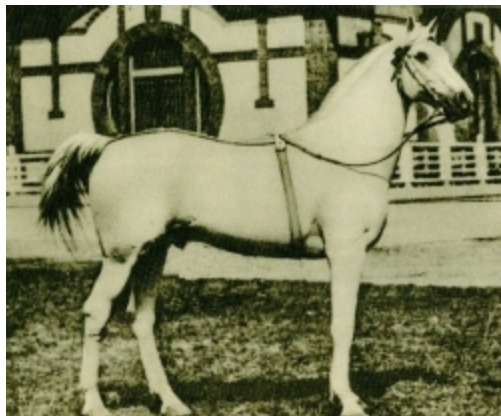
Actualmente y a pesar de la gran cantidad de importaciones producidas, se conocen 4 líneas paternas directas en el Pura Raza Árabe Español (Fenaux, 1995):

- *Seanderich* importado por Yeguada Militar en 1908, con 3 ó 4 sublíneas.
- *Ursus* importado por Yeguada Militar en 1912.
- *Razada* importado por el Duque de Veragua en 1926.
- *Bagdad* importado por Yeguada Militar en 1925.

Además de esas 4 líneas paternas, el semental *Wan dick* importado por Yeguada Militar en 1908, aunque no es línea paterna directa, es considerado un semental de una gran importancia.

Las familias maternas consideradas están mucho más diversificadas y se clasifican según su origen:

- 13 proceden del desierto: *Sambry, Farja I, Ghazel, Ymm, Fatima, Bint, Selimieh, Beni-Kaled, Saklauia-Yadran...*entre otras.
- 4 de Polonia: *Kadranka, Kebrebassa, Damietta, Reyna.*
- 1 de Francia: *Sultane.*
- 1 de Inglaterra.
- 15 son familias procedentes de la ganadería del Duque de Veragua: *Veracruz, Verana, Veranda, Verapaz, Veralca...*entre otras.



Wan dick, semental PRá importado de Polonia por la Yeguada Militar. Reproducción de 1950 (Yeguada Militar de Córdoba).

3.1.2. Censos y localización geográfica. La cría en España

El caballo de PRá es una raza clasificada como "Española" según el R.D. 1682/1997 por el que se publica el Catálogo Oficial de Razas de Ganado de España y sus modificaciones posteriores (Orden APA 2420/2003 y 661/2006).

Es la tercera raza equina española en cuanto a censo se refiere, con un total de 8.949 ejemplares según fuentes del MAPyA (2003). Se encuentra extendida por todo el país, pero es en Andalucía donde mayor número de cabezas hay censadas constituyendo casi el 36% del total, seguida de Cataluña con un 13% del efectivo nacional. Esta distribución se debe a la evolución histórica de la raza en España. Durante años los criadores de árabe estuvieron asentados en el sur y centro de la Península. A partir de los 70 se extendieron a las Baleares, el Norte de España y a Cataluña.

En la Figura 1 se muestra la evolución de los animales inscritos en Libro Genealógico de la raza desde su fundación hasta el año 2004. Se puede observar un ascenso más o menos progresivo de los animales nacidos hasta el año 2004, observándose una explosión demográfica a partir de los años 70. El descenso observado durante los años 1935-44 es producido por los hechos acaecidos durante la Guerra Civil Española. La proporción de machos y hembras se ha mantenido más o menos constante a lo largo de las décadas siendo más numerosas las yeguas.

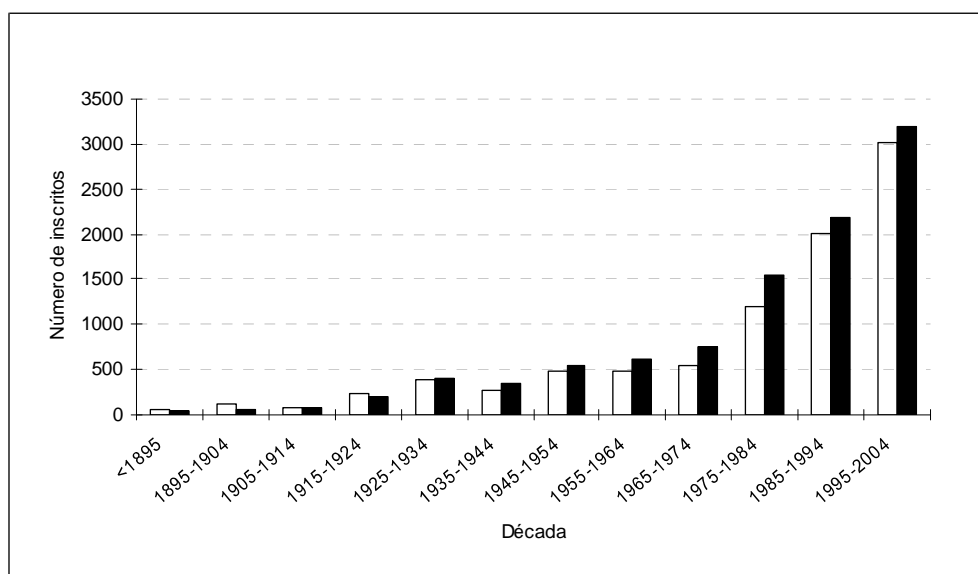


Figura 1. Evolución de los nacimientos de los animales inscritos en el Libro genealógico del caballo de Pura Raza Árabe (yeguas en barras negras y machos en barras blancas). (Fuente: elaboración propia a partir de datos del Libro Genealógico Español de caballo de Pura Raza Árabe).

Desde el punto de vista de las ganaderías, el número total que se han inscrito a lo largo de la historia del Libro Genealógico del Caballo de PRÁ es de 1.492. De ellas sólo 620, el 41,6 % del total, han inscrito en algún momento animales que después han actuado como reproductores (yeguas o sementales).

En cuanto a la estructura de la población por ganaderías, éstas pueden ser clasificadas según su estrategia de cría (Vasallo *et al.*, 1986):

- Ganaderías Núcleo: Son aquellas en las que el criador utiliza únicamente sus propios sementales en la cría, no usan los sementales de otras ganaderías aunque sí venden reproductores.
- Ganaderías Multiplicadoras: En ellas el criador además de utilizar sementales de otras ganaderías pone a la venta los propios.
- Ganaderías Comerciales: En estas ganaderías únicamente se utilizan sementales externos y nunca se venden los que nacen en ellas. Se abastecen de la genética de otras ganaderías sin ofrecer sus productos a los demás.

Dentro de las ganaderías inscritas de PRá sólo se han encontrado ganaderías de tipo multiplicador y comercial, siendo la mayoría de las ganaderías que poseen un alto porcentaje de animales inscritos de tipo multiplicador (56,8%).

3.1.3. Descripción morfológica. Prototipo racial

La “World Arabian Horse Organization” (Organización Mundial de Caballos Árabes, WAHO) es el organismo internacional que establece las pautas de cría y registro en los Libros Genealógicos. Fue establecida en los años 70. En 1999 la WAHO definió al caballo de Pura Raza Árabe como aquel caballo que aparece inscrito en un Libro Genealógico de la raza aceptado como válido por esta organización. No existe en España una normativa específica que describa su patrón racial, pero son muchas las descripciones que se han realizado sobre esta raza.

Tradicionalmente se han descrito 3 líneas principales que dan lugar a un determinado tipo de caballo Árabe: Kuhaylan (fuerza), Saqlai (belleza) y Muniqui (velocidad). Hoy día no persisten en pureza, pero aún existen ejemplares que podrían englobarse en alguna de ellas.

Fenau en 1995 hizo la siguiente descripción morfológica basándose en una extensa revisión bibliográfica:

- Caballo de tipo mesomorfo.
- La cabeza presenta un perfil recto o ligeramente cóncavo y es ancha. El hocico es pequeño y los ollares grandes. Sus ojos son grandes, brillantes y expresivos, situados por delante de la línea del frontal. Las orejas son moderadamente cortas.
- El cuello es razonadamente largo y proporcionado al tamaño del caballo y de su cabeza.
- El tronco es cuadrado, nunca más largo que alto, con buena cruz, caja torácica amplia y dorso y lomo cortos. La cola es de nacimiento alto.
- Las extremidades anteriores presentan una espalda larga y bien inclinada para favorecer las extensiones. Los aplomos son rectos con buenas

articulaciones, limpias y fuertes. El tercio posterior es bien musculado y fuerte.

- Destaca por su resistencia, elegancia y energía. Es un animal de temperamento manejable, noble, potente, ágil y longevo.



Titánic, semental PRá, propiedad de D. Nicolás Domecq. Fotografía tomada por grupo MERAGEM.



Agadir, semental PRá, propiedad de Yeguada Militar. Fotografía tomada por grupo MERAGEM.

3.1.4. Gestión y utilidad de la raza

El Plan de Ordenación Zootécnica y Fomento del Sector equino del entonces denominado Ministerio de Agricultura, Pesca y Alimentación iniciado en 2002, cuyo objetivo es la mejora de cada una de las razas equinas explotadas en España, ha servido de impulso para el inicio de la mejora genética del PRá en España.

En España, la asociación nacional encargada de la gestión de la raza es La Asociación Española de Criadores de Caballos Árabes (AECCA), fundada en 1970 al mismo tiempo de la creación de la WAHO, engloba actualmente a un total de 105 socios ganaderos.

La llevanza del Libro Genealógico del PRá (inscripción de nuevos individuos) ha sido realizada por el Ministerio de Defensa a través del Fondo de Explotación de los Servicios de Cría Caballar y Remonta hasta el año 2008, año en que sus competencias han sido traspasadas para ser gestionado por parte de la propia Asociación. Este Libro Genealógico, cumpliendo el Real Decreto 662/2007, consta de un Registro de Nacimientos y un Registro Principal. En cuanto a los sistemas reproductivos, la WAHO acepta tanto la inseminación artificial como la transferencia de embriones, existiendo algunas limitaciones para ambas:

- Para la inseminación artificial se necesita un permiso para cada semental del que se quiera utilizar semen fresco, refrigerado o congelado. El animal debe estar genotipado al igual que las yeguas inseminadas artificialmente con objeto de asegurar un correcto control de paternidad. Con cada inseminación deberá realizarse un informe donde se indiquen los datos identificativos, tanto del semental, como de la yegua, de sus propietarios y

de la fecha de colección del semen. Después de la muerte del animal su semen podrá ser utilizando sin límite de tiempo. Todos los animales nacidos por inseminación artificial tendrán que someterse a un control de paternidad antes de ser registrados en el Libro Genealógico.

- Para la transferencia de embriones, al igual que los sementales usados para colección de semen, las yeguas donantes deben obtener el permiso correspondiente, debiendo estar genotipadas tanto éstas como las receptoras. A los animales nacidos por transferencia embrionaria se les tomará muestra de pelo y de sangre para realizar el control de paternidad antes de su inscripción. Se establece una restricción de un potro al año (salvo parto gemelar) para el registro de hijos procedentes de yeguas donantes (Valera *et al.*, 2007).

La explotación de la raza viene determinada por su versatilidad como caballo de silla pudiendo destacar en las diferentes disciplinas ecuestres (Raid, Concurso Completo de Equitación, Doma...). Así, son destacables sus actuaciones en las carreras de resistencia (Raid).

El Raid es una disciplina en la que se pone de manifiesto la resistencia del caballo y la estrategia del jinete para poder realizar trayectos de larga distancia (pueden alcanzar los 200 Km) en espacios naturales con diferentes tipos de recorrido. La prueba se realiza a contrarreloj, manteniendo la velocidad adecuada y consiguiendo pasar los controles veterinarios realizados durante el recorrido.



Detalle de una competición de Raid, Torrijos (Toledo). Final Ciclos de Caballos Jóvenes 2008.



Detalle del control veterinario en un Raid. Auscultación abdominal.

Las carreras de Raid son organizadas por la Real Federación Hípica Española (RFHE), aunque desde 2006 ha intervenido la AECCA en la organización de

diferentes Raid orientados como Pruebas de Selección de Caballos Jóvenes en colaboración con el Ministerio de Medio Ambiente y Medio Rural y Marino (Valera *et al.*, 2007).

En la Tabla 1 se indican las principales categorías nacionales e internacionales de Raid existentes hasta el momento y en las que pueden inscribirse equinos de todas las razas.

Tabla 1. Categorías de los concursos de Raid a nivel Nacional e Internacional (Valera *et al.*, 2007).

Concursos Nacionales		Concursos Internacionales	
Categorías	Distancia (km)	Categorías	Distancia (km)
Promoción	20-60	CEI*	80-119
CEN*	60-99	CEI**	100-160
CEN**	100-160	CEI***	140-160
CEN***	140-200	CEI****	160-200
CEN ****	160-200		

Desde el punto de vista morfológico, es la ECAHO (European Conference of Arab Horse Organization) la entidad que establece los reglamentos referentes a los concursos morfológicos homologados para la raza, tanto dentro como fuera de España.

Dependiendo del número de Secciones o Clases de la categoría del concurso y del número de animales participantes, existen diferentes categorías de concursos: A, B, C, D, Campeonato de Europa, Copa de las Naciones y Campeonato del Mundo.

3.1.5. Programa de Mejora

El caballo de PRá tiene aprobado por Resolución de la Dirección General de Ganadería con fecha 1 de septiembre de 2005 un Esquema de Selección elaborado por el Grupo PAI de investigación AGR-158 (MERAGEM) a petición de la Asociación de Criadores de Caballos de Pura Raza Árabe (AECCA) y siguiendo la normativa (Orden APA 1018/2003).

Dentro de sus objetivos se establece la consecución de un animal, que además de mantener las características raciales del caballo Árabe, sea noble, bello, inteligente, correcto y resistente, con temperamento, habilidades y movimientos adecuados para destacar en la disciplina ecuestre en la que participe, tanto a nivel nacional como internacional.

El Esquema de Selección del PRá se ha estructurado en 5 fases según exige la Orden APA/1018/2003 (Figura 2).

1ª Fase. Registro de nacimientos. Libro de Registro

Durante esta primera fase del Esquema se realizará el reconocimiento e identificación de los nuevos productos, así como el control de filiación mediante las pruebas de ADN. Los animales serán inscritos en el correspondiente Libro Genealógico de la Raza.

En esta fase se podrá realizar un control genético de aquellos defectos con transmisión hereditaria que se considere oportuno, eliminando del Esquema a aquellos animales portadores de caracteres deletéreos o causantes de enfermedades, con anomalías congénitas o con alteraciones cromosómicas transmisibles. Estos estudios podrán ser complementados con un análisis inmunogenético y citogenético, si la naturaleza del problema así lo requiere.

2ª Fase. Valoración genealógica

A través de la información recogida en el correspondiente Libro Genealógico de la raza se realizará el análisis genealógico de los animales. Se determinará para cada uno de los animales adscritos al Esquema de Selección su nivel de consanguinidad, la influencia genética de las líneas principales de la raza y el flujo de genes entre los participantes en el Esquema. Asimismo, se determinará el sistema de conexiones entre las distintas ganaderías del Esquema, lo que permitirá establecer, junto con los posteriores resultados de la valoración genética, el sistema de cruzamientos entre los individuos.



Morata, yegua PRá, propiedad de D. Juan Jódar. Fotografía tomada por el Grupo MERAGEM.

3ª Fase. Control de rendimientos y valoración genética individual: Joven Reproductor Recomendado

Durante esta fase del Esquema de Selección se realizará el control de rendimientos para la morfología, los movimientos, el temperamento (en pruebas de campo específicas) y la aptitud funcional para el raid en Ciclos de Caballos Jóvenes y Competiciones de Raid, tanto nacionales como internacionales, que por sus características se consideran homologables. Una vez procesados y analizados genéticamente los datos, se determinarán los “Jóvenes Reproductores Recomendados”.

Se establecerán dos categorías:

- Joven Reproductor Recomendado para Morfología y Movimientos: animales de entre 4 y 7 años con resultados propios que tengan un valor genético superior a la media.

- Joven Reproductor Recomendado para la disciplina de Raid, animales de entre 4 y 7 años con resultados propios que tengan un valor genético superior a la media.

4ª Fase. Valoración genética por la descendencia: Reproductor Mejorante y de Élite

Los reproductores con 7 o más años, y muy especialmente los jóvenes reproductores recomendados en la 3ª fase del esquema, y que cuenten con suficiente número de animales emparentados genealógicamente con resultados de control de rendimientos, serán valorados genéticamente a través de su descendencia. Se establecerán igualmente dos categorías:

- Reproductor Mejorante para Morfología y Movimientos: animales mayores de 7 años con un índice genético superior a la media y una fiabilidad superior al 50%.
- Reproductor Mejorante para la disciplina de Raid, animales mayores de 7 años con un índice genético superior a la media y una fiabilidad superior al 40%.



Punto, semental PRÁ, propiedad de D. Gonzalo Moreno. Fotografía tomada por el Grupo MERAGEM.

Para que un animal sea reconocido como Reproductor de Élite, es necesario que sea Mejorante en más de una de las categorías establecidas para esta raza y haya destacado por sus propios méritos deportivos. A pesar de que para ser incorporado en el registro de Reproductor Élite es preferible que con anterioridad el reproductor hubiese conseguido la calificación de Joven Reproductor Recomendado, un animal puede ser considerado como Reproductor Élite y consecuentemente Mejorante para alguna o varias de las anteriores disciplinas, única y exclusivamente a través de la valoración genética de los animales con los que guarda una relación de parentesco.

5ª Fase. Catálogo de Reproductores

Los ganaderos y criadores de Caballos de Pura Raza Árabe dispondrán de forma periódica de un Catálogo de Reproductores, donde se incluirán los Reproductores Élite y los Reproductores Mejorantes (tanto yeguas como sementales). Se incluirá también la relación de los Jóvenes Reproductores Recomendados con el objeto de incrementar el número de descendientes sobre los que poder realizar su futura valoración genética para ser incorporados como Reproductores Mejorantes.

La valoración para la aptitud morfológica será realizada mediante la Calificación Morfológica Lineal (CML). Esta metodología permite obtener información objetiva y fiable con la que estimar parámetros en índices genéticos de elevada fiabilidad (Gómez *et al.*, 2006). Divide al animal en rasgos anatómicos simples y el calificador describe cada uno de esos rasgos. Así, se puede determinar el modo de transmisión de los caracteres de interés y permite corregir o mejorar los animales de una ganadería dependiendo de sus necesidades específicas. La CML incluye caracteres correlacionados con la funcionalidad de la raza. Por ello, la mejora que será llevada a cabo mediante el uso de esta metodología será morfo-funcional.

En cuanto a los aspectos funcionales, por el momento es la disciplina de Raid la que está sometida a valoración genética por ser la que tiene mayor número de caballos de PRá participantes.

En el año 2006 se realizó la primera valoración genética para la disciplina de raid en el caballo de PRá para lo cual se ha contado con datos de 249 caballos con un total de 547 participaciones en raids de diferentes categorías.

La valoración genética se ha realizado para los caracteres “puesto clasificatorio” y “tiempo de carrera”. Los factores sistemáticos que se incluyeron en el modelo de valoración fueron el año de celebración de la prueba, la zona geográfica donde se realiza la prueba y los kilómetros del recorrido. Además, se incluyeron como covariables el número total de participantes en la prueba para el carácter “puesto clasificatorio” y el tiempo medio de carrera para el carácter “tiempo de carrera”.

Las heredabilidades obtenidas presentan un valor medio-bajo: 0'18 para el puesto clasificatorio y 0'13 para el tiempo (Cervantes *et al.*, 2006). La evolución del valor genético para dichos caracteres mostró que el progreso genético ha sido escaso hasta el momento, pero la elevada variabilidad del carácter asegura un progreso adecuado si se aplica una apropiada intensidad de selección para dichos caracteres.

El Esquema del PRá queda abierto para incorporar valoraciones genéticas de otras aptitudes funcionales (enganche, salto, doma, carrera, C.C.E., etc) en el momento que exista un número suficiente de animales con resultados de pruebas deportivas de la categoría.



Estel, macho PRá, en ejercicio de cross de una competición de Concurso Completo de Equitación. Fotografía cedida por D. Albert Hermoso.

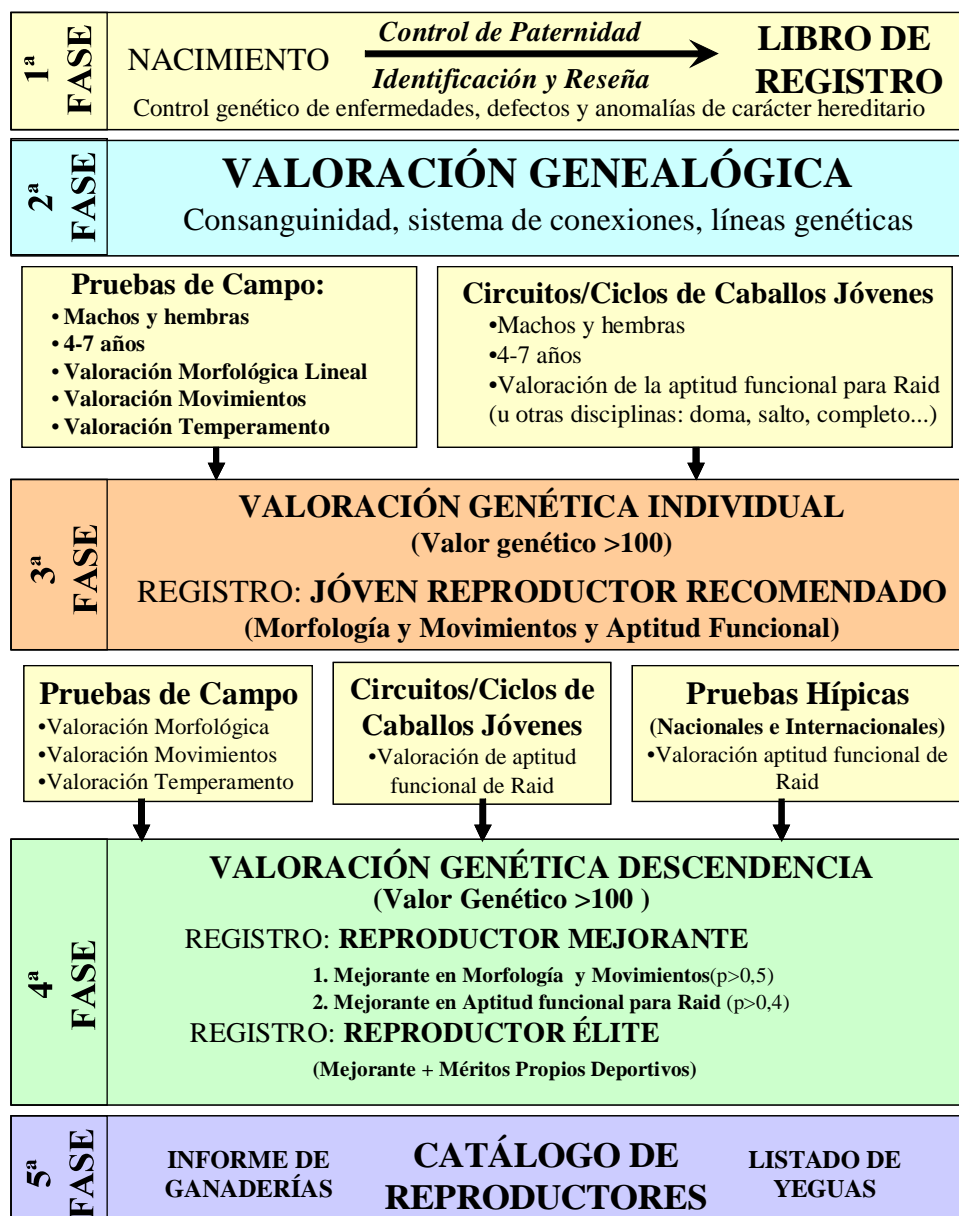


Figura 2. Diagrama del Esquema de Selección del caballo de Pura Raza Árabe (Fuente: <http://www.marm.es/>).

3.2. RAZAS DERIVADAS DEL ÁRABE

Como se ha dicho anteriormente el caballo de PRá, dada sus características morfológicas y funcionales participa en la formación de diversas razas. En España por su importancia destacan tres: el caballo Anglo-Árabe (A-á), el caballo Hispano-Árabe (H-á) y el Caballo de Deporte Español (CDE).

3.2.1. La raza Anglo-Árabe

La raza A-á está clasificada como raza equina “Española” según el Catálogo oficial de razas de Ganado de España (R.D. 1682/1997 y modificaciones posteriores). Esta raza, procedente de Francia, se originó por el cruce de dos sementales árabes y tres yeguas inglesas en las yeguas Pompadour, Lespinats y Eugène Gayot, en 1843 y se expandió posteriormente al resto del mundo (Llamas *et al.*, 1992).

En España su cría comenzó muy pronto por parte del Ministerio de Defensa. De hecho, los primeros caballos Pura Raza Árabe que se importaron fueron con el objetivo de obtener animales cruzados buscando las buenas aptitudes de la raza (Maxwell, 1995). La publicación del primer tomo del Registro-Matricula de caballos y yeguas de caballos Anglo-Árabe (A-á) en España data de 1884.

Ocupa el quinto lugar en las razas equinas en cuanto a censo se refiere, existiendo, según fuentes del MAPyA (2003), un total de 3.421 cabezas. En Andalucía es donde se concentra la mayor parte de la cabaña ganadera de A-á, constituyendo el 57% del censo total de la raza. La Orden APA/42/2003 establece las normas zootécnicas que regulan la gestión de la raza y establece las secciones presentes en su Libro Genealógico.

El propósito de la creación de esta raza, y que sigue teniendo vigencia actualmente, consiste en producir caballos deportivos de alta calidad que destaquen en la mayoría de las disciplinas ecuestres.

3.2.2. La raza Hispano-Árabe

El cruce de individuos de Pura Raza Español y de Pura Raza Árabe para la creación de animales H-á comenzó a realizarse también de forma temprana en el siglo XX (Steen, 2006), aunque el Libro Registro Matricula de la raza no se creó hasta 1986 (Orden 70/1986, de 21 de agosto del Ministerio de Defensa).

Está considerada como raza “Autóctona de Protección Especial” según el Catálogo oficial de razas de Ganado de España (R.D. 1682/1997 y modificaciones posteriores). Su censo es de 2.484 cabezas, situándose un 37% del total en Andalucía (MAPyA, 2003). Las normas zootécnicas de la raza están publicadas en la Orden APA 3277/2002. Actualmente su objetivo de cría está orientado a actividades como la Doma Vaquera, el Acoso y Derribo y el TREC (Técnicas de rutas ecuestres de competición).

3.2.3. El Caballo de Deporte Español

Por último el CDE se trata de una raza de reciente creación, como lo demuestra el hecho de que su Registro Fundacional se cerró en Diciembre de 2004. Está formada por todos aquellos animales de dos y tres “sangres”, competitivos en las diferentes disciplinas hípicas, que no podían competir en eventos de carácter internacional por no estar inscritos en ningún Libro Genealógico.

Según datos del MAPyA (2003), la raza de CDE posee un censo de 2.350 animales. Está distribuida por todas las regiones de España, en Andalucía se concentra la mayor parte de la población, contando con más de un 21% del censo total.

La primera normativa propia de la raza, se creó en España en 2002, mediante la Orden APA/3318/2002, del 23 de Diciembre, por la que se establecen las normas zootécnicas del CDE. Según ésta, podrá considerarse como CDE a todos aquellos animales que, procediendo de alguno de los cruces recogidos en la tabla 2, hayan nacido a partir de 1998 o a partir de 1992, en cuyo caso, deben disponer de identificación propia y de los padres que deben estar vivos o identificados por marcadores genéticos.

Tabla 2. Cruces admisibles para que el producto sea inscrito en el Libro Genealógico del CDE.

	PSI	PRá	A-á	H-á	PRE	TE	MALL	MEN	CDEx.	CDE
PSI	-	-	-	CDE	CDE	CDE	CDE	CDE	CDE	CDE
PRá	-	-	-	-	-	CDE	CDE	CDE	CDE	CDE
A-á	-	-	-	CDE	CDE	CDE	CDE	CDE	CDE	CDE
H-á	CDE	-	CDE	-	-	CDE	CDE	CDE	CDE	CDE
PRE	CDE	-	CDE	-	-	CDE	CDE	CDE	CDE	CDE
TE	CDE	CDE	CDE	CDE	CDE	-	CDE	CDE	CDE	CDE
MALL	CDE	CDE	CDE	CDE	CDE	CDE	-	CDE	CDE	CDE
MEN	CDE	CDE	CDE	CDE	CDE	CDE	CDE	-	CDE	CDE
CDEx	CDE	CDE	CDE	CDE	CDE	CDE	CDE	CDE	CDE*	CDE
CDE	CDE	CDE	CDE	CDE	CDE	CDE	CDE	CDE	CDE	CDE

PSI: Pura Sangre Inglés; PRá: Pura Raza Árabe; A-á: Anglo-árabe; H-á: Hispano-árabe; PRE: Pura Raza Española; MALL: Mallorquín; MEN: Menorquín; TE: Trotador Español; CDEx: Caballo de Origen Extranjero inscrito en Libros Genealógicos reconocidos por la World Breeding Federation for Sport Horses (WBFSH).

* Los productos de estos cruces darán lugar a CDE si lo solicita el criador, excepto si esos animales están inscritos en otro Libro Genealógico oficialmente reconocido.

Con la elaboración y puesta en marcha del Esquema de Selección de la raza se persigue obtener animales con altos rendimientos deportivos en las competiciones en las que participen. Hasta el momento son tres las disciplinas en las que se establece selección para los individuos de la raza: el Salto de Obstáculos, la Doma Clásica y el Concurso Completo de Equitación (Valera *et al.*, 2007).

3.3. ESTRUCTURA GENÉTICA DE LA POBLACIÓN

El estudio de la estructura genética de una población permite conocer cómo se ha llevado a cabo el flujo de genes en la misma, proporcionando información sobre el número de genes que la ha originado y estimando en qué medida participan en la población actual. Este conocimiento proporciona información sobre su grado de variabilidad genética y permite tomar decisiones en relación a su gestión genética.

El estudio de la estructura genética de una población mediante análisis genealógico ha sido realizado en la mayoría de las especies, como el bovino de carne y leche y ovino entre otras (Boichard *et al.*, 1997; Sölkner *et al.*, 1998; Gutiérrez *et al.*, 2003; Goyache *et al.*, 2003). En cuanto a la especie equina encontramos en la bibliografía estudios poblacionales basados en la información del pedigrí, tanto en razas extranjeras como españolas. Concretamente en la raza Árabe, Moureaux *et al.*, (1996) realizó este análisis en 4 razas francesas, incluyendo al caballo de Pura Raza Árabe Francés y más recientemente en 2004, Głazewska y Jezierski, llevaron a cabo la caracterización del caballo de Pura Raza Árabe polaco. Destacan también los estudios genealógicos en el caballo Lipizzano (Zechner *et al.*, 2002) y el Hannoveriano (Hamann y Distl, 2008). En cuanto a razas españolas, hay que reseñar el realizado por Valera *et al.* (2005) sobre el caballo de Pura Raza Español y la estirpe Cartujana.

A continuación se describen una serie de de parámetros que caracterizaran la estructura genética de una población.

3.3.1. Intervalo generacional

El estudio del intervalo generacional en el análisis de la estructura genética de poblaciones es, en sí mismo, importante ya que las pérdidas de variabilidad genética se producen por generación. Así, si el intervalo generacional es corto las pérdidas de variabilidad genética se producen a mayor velocidad por unidad de tiempo. Pero además, el intervalo generacional es un parámetro que permite traducir incrementos de otros parámetros por unidad de tiempo en incrementos de los mismos por generación.

Este parámetro se define como la edad media de un animal reproductor a la que es reemplazado en esta función por un descendiente suyo y se calcula como la edad media de los padres cuando nacen sus descendientes que luego serán reproductores (James, 1977). Se puede obtener para las 4 vías posibles (padre-hijo, padre-hija, madre-hijo y madre-hija).

3.3.2. Grado de profundidad de pedigrí

Dado que en poblaciones cerradas las pérdidas de variabilidad genética se acumulan por generación, los parámetros genealógicos que la miden deben ser siempre referidos a la cantidad de información genealógica existente, por lo que su medida resulta imprescindible (Boichard *et al.*, 1997; Gutiérrez *et al.*, 2003).

La existencia de solapamiento generacional y el desigual conocimiento genealógico por las vías paterna y materna imposibilita la asignación de individuos

a generaciones discretas. Entre considerar que un individuo pertenece a la generación g si se conocen sus 2^g antepasados (generaciones completas), o considerar g como el número de generaciones que separan al individuo de su antecesor más lejano (generaciones máximas) existen muchas posibilidades. Una forma de afrontar esta dificultad es utilizar el número de generaciones completas equivalentes que se obtiene como la suma para todos los antepasados conocidos de $(\frac{1}{2})^n$ siendo n el número de generaciones que separan al individuo de cada antepasado (Maignel *et al.*, 1996).

También es muy utilizado el índice de profundidad de pedigree de MacCluer *et al.* (1983), que proporciona el porcentaje de padres, abuelos, etc., o una representación gráfica del porcentaje de antepasados conocidos en las generaciones más recientes distinguiendo las vías paternas y maternas (Gutiérrez *et al.*, 2003; Ron Garrido *et al.*, 2006).

3.3.3. Coeficiente de Consanguinidad (F)

El coeficiente de consanguinidad, coeficiente de endogamia o identidad por descendencia de un individuo se define como la probabilidad de que los dos alelos que lleva un individuo en cualquier locus sean idénticos por descendencia (Wright, 1922; Malécot, 1948). Es probablemente el parámetro más importante en el análisis de la estructura genética ya que mide la probabilidad de homocigosis en una genealogía, y es precisamente la homocigosis la causante de la depresión consanguínea y de la manifestación de genes deletéreos.

Así, los niveles de consanguinidad altos conducen a fenómenos de depresión consanguínea. De forma general, la depresión produce la aparición de efectos deletéreos en la población y una disminución del rendimiento de los caracteres muy afectados por la acción no aditiva de los genes, es decir de los caracteres de baja heredabilidad (Ollivier, 1981), que están generalmente relacionados con la reproducción (Wright, 1977; Pirchner, 1985).

En la especie equina se han descrito fenómenos de depresión consanguínea que afectan a la tasa de crecimiento y la conformación (Radomska *et al.*, 1984; Gandini *et al.*, 1992; Gómez *et al.*, 2008), a la capacidad reproductiva (Weitkamp *et al.*, 1982; Klemetsdal *et al.*, 1989; Cunningham, 1991; Oom, 1992; Valera *et al.*, 1996; Van Eldik *et al.*, 2006) y a la funcionalidad (Barreiros y Barata, 1982 y Wilkens *et al.*, 1990).

3.3.4. Coeficiente de Relación Media (AR)

El coeficiente Relación Media (AR) de un individuo se define como la probabilidad de que un alelo escogido al azar de la población pertenezca a ese individuo. AR puede interpretarse como la representación del individuo en el pedigrí independientemente de la generación en que se encuentre. Por tanto es una medida del porcentaje que por término medio comparte cada individuo con el resto de los individuos en el pedigrí. El cálculo de AR se basa en la media de los elementos de cada fila de la matriz de relaciones aditivas, lo que daría lugar a un vector $\mathbf{c}' = (1/n)\mathbf{1}'\mathbf{A}$, siendo \mathbf{A} la matriz de relaciones aditivas de tamaño $n \times n$. El

algoritmo utilizado para obtener estos valores se basa en el desarrollado por Quaas (1976) para construir la propia matriz **A**. Conviene recordar que los coeficientes de relación aditiva son el doble de los coeficientes de coascendencia, y que el coeficiente de coascendencia entre dos individuos es igual a la consanguinidad de un descendiente de ambos.

Las ventajas de utilizar *AR* son:

- Su coste computacional es similar al de la construcción de **A** ya que comparten el algoritmo.
- *AR* de un fundador indica su porcentaje de contribución genética a la población, es decir el porcentaje de población originado por él.
- Los coeficientes *AR* pueden utilizarse como medida de consanguinidad de la población globalmente al tener en cuenta simultáneamente la consanguinidad y la coascendencia.
- *AR* puede ser utilizado como índice para mantener el stock inicial mediante la utilización preferente como padres de individuos con el menor *AR*.
- *AR*, puede utilizarse para predecir la consanguinidad a largo plazo de una población, porque tiene en cuenta el porcentaje del pedigrí completo originado por cada fundador. En definitiva, *AR* puede utilizarse para obtener el tamaño efectivo de la población fundadora mediante el inverso de la suma de los cuadrados de los valores de los fundadores.

3.3.5. Probabilidad de origen de los genes

Todos los genes presentes en una población proceden de alguno de sus fundadores, pero la representación de cada uno de ellos, en los individuos tomados como referencia es muy variable dependiendo del uso preferente o no que se haga de cada uno de los individuos reproductores. Así, la probabilidad de origen de los genes proporciona una información de gran valor en el conocimiento del flujo de los genes y, en consecuencia, de la estructura genética de la población.

Así, para medir la variabilidad genética y conocer las consecuencias de la política de apareamientos llevada a cabo en una población, una de las informaciones más interesantes procede del análisis de la contribución genética en términos de animales fundadores y de ancestros.

Dado que fundador y ancestro son conceptos que se confunden habitualmente, en esta memoria serán entendidos de la siguiente manera:

Fundador: es aquel individuo con padre y madre desconocidos. Cuando un individuo tiene un único parental desconocido, éste se considera fundador fantasma.

Ancestro: es aquel individuo que ha influido más que sus ascendientes en la variabilidad genética presente en la población. Un ancestro puede ser también fundador.

Basándose en la desigual probabilidad de origen de los genes por fundadores o ancestros se definen tres parámetros relacionados, el tamaño efectivo de fundadores, el tamaño efectivo de ancestros y el número efectivo de genomas fundadores. El primero nunca puede ser inferior al segundo ni éste al tercero.

3.3.5.1. Número efectivo de fundadores (f_e)

Se define como el número de fundadores que contribuirían de forma equilibrada para producir la misma diversidad genética que la población estudiada. Se calcula mediante:

$$f_e = \frac{1}{\sum_{k=1}^f q_k^2} \quad [1]$$

donde q_k es la probabilidad de origen de los genes del fundador k (James, 1972; Lacy, 1989). Si se considera todos los individuos del pedigrí como referencia, se puede obtener utilizando los coeficientes AR de los fundadores.

3.3.5.2. Número efectivo de ancestros (f_a)

El parámetro f_a es el mínimo número de ancestros, no necesariamente fundadores, que explicarían la diversidad genética de la población y complementa la información proporcionada por el parámetro anterior, dado que tiene en cuenta la pérdida de variabilidad genética que se debe a cuellos de botella (Boichard *et al.*, 1997). Se calcula de manera muy similar al anterior:

$$f_a = \frac{1}{\sum_{j=1}^a q_j^2} \quad [2]$$

En este caso q_j es la contribución marginal de un ancestro j , que es la contribución genética realizada por un ancestro que no es explicada por otros ancestros elegidos previamente.

3.3.5.3. Número efectivo de genomas fundadores (f_g)

Es definido como el número de fundadores que producirían la diversidad genética presente en la población en estudio si todos los fundadores estuvieran igualmente representados y no se hubiera producido pérdida de alelos (Ballou y Lacy, 1995). En este sentido este parámetro tiene en cuenta todas las causas posibles de las pérdidas de variabilidad genética y será el menor valor de los tres. Se puede calcular según Caballero y Toro (2000), mediante el doble de la inversa de la coascendencia media de los individuos incluidos en la población.

3.3.6. Estadísticos F de Wright

Mediante estos parámetros, F_{IT} , F_{ST} y F_{IS} se puede estudiar la estructura poblacional (Wright, 1978), suponiendo la existencia de una metapoblación constituida por varias subpoblaciones, que pueden ser definidas en función de distintos sexos, áreas, ganaderías, etc. Se pueden explicar en términos de heterosis y en términos de consanguinidad:

F_{ST} : es el coeficiente de consanguinidad medio (o reducción de la heterosis) de las subpoblaciones respecto de la metapoblación. El valor F_{ST} se utiliza como distancia entre las diferentes subpoblaciones, su valor está entre 0 y 1 y cuanto más alto es, más diferenciación genética hay entre las subpoblaciones

F_{IS} : es el coeficiente de consanguinidad medio (o reducción de la heterosis) de un individuo respecto a su subpoblación. El estadístico F_{IS} expresa la presencia de apareamiento no aleatorio dentro de la subpoblación. Cuando $F_{IS} > 0$, se están produciendo apareamientos entre individuos relacionados. Cuando $F_{IS} < 0$, la frecuencia de apareamientos entre individuos no relacionados es mayor (se produce un exceso de heterocigotos)

F_{IT} : es el coeficiente de consanguinidad medio (o reducción de la heterosis) de individuo respecto de la metapoblación.

Caballero y Toro (2000, 2002) han formalizado las herramientas de pedigrí necesarias para el análisis de la diferenciación genética en poblaciones subdivididas a partir de los coeficientes de coascendencia entre los individuos de dos subpoblaciones, i y j , de una metapoblación, incluyendo los $N_i \times N_j$ pares posibles. Para una población i , la coascendencia media, la autocoascendencia de los N_i individuos y la consanguinidad media serían respectivamente, f_{ii} , s_i y $F_i = 2s_i - 1$. La distancia media entre individuos de dos subpoblaciones i y j

sería $D_{ij} = \left[(s_i + s_j) / 2 \right] - f_{ij}$. Esta distancia se ha llamado *kinship distance* (D_k)

por Eding y Meuwissen (2001) cuando se trabaja con datos moleculares. De estos parámetros y las correspondientes medias de la metapoblación Caballero y Toro (2000, 2002) obtienen la distancia genética entre dos subpoblaciones i and j (distancia mínima de Nei; Nei, 1987) como

$D_{ij} = D_{ij} - \left[(D_{ii} + D_{jj}) / 2 \right] = \left[(f_{ii} + f_{jj}) / 2 \right] - f_{ij}$, y su media global como

$\bar{D} = \frac{\sum_{i,j=1}^n D_{ij} N_i N_j}{N_T^2}$. Finalmente los estadísticos F de Wright (1978) se obtienen

como:

$$F_{IT} = \frac{\bar{F} - \bar{f}}{1 - \bar{f}} \quad [3] \quad F_{ST} = \frac{\bar{f} - \bar{f}}{1 - \bar{f}} = \frac{\bar{D}}{1 - \bar{f}} \quad [4] \quad F_{IS} = \frac{\bar{F} - \bar{f}}{1 - \bar{f}} \quad [5]$$

donde \bar{f} y \bar{P} son, respectivamente, la coascendencia y la consanguinidad medias de la subpoblación considerada, y \bar{f} la coascendencia media de la metapoblación, de modo que todos estos parámetros se relacionan de la siguiente manera:

$$(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST}) \quad [6]$$

3.4. CÁLCULO DEL CENSO O TAMAÑO EFECTIVO

3.4.1. El valor relativo del coeficiente individual de consanguinidad

El coeficiente de consanguinidad proporciona la probabilidad de homocigosis de un gen en el individuo en función de su genealogía, lo que resulta útil desde diversos puntos de vista:

- Los valores medios pueden ser utilizados para informar del grado de variabilidad genética de la población, proporcionando una herramienta útil en la descripción de la estructura genética de poblaciones Wright (1978).
- Dado que la homocigosis conduce a la manifestación de genes letales o subletales (Cavalli-Sforza y Bodmer, 1971), estos coeficientes proporcionan una medida de probabilidad de manifestación de este tipo de caracteres; aunque teóricamente esta afirmación no ofrece dudas, esto no es estrictamente cierto en la práctica, ya que, aunque dispusiéramos de la genealogía completa del individuo en un elevado número de generaciones, la selección natural opera en contra de la homocigosis.
- Finalmente, la utilización de estos coeficientes en relación con variables cuantitativas permitiría la estimación de la depresión consanguínea que se produciría principalmente en caracteres reproductivos (Van Eldik *et al.*, 2006; González-Recio *et al.*, 2007), pero también en otro tipo de caracteres (Wilkens *et al.*, 1990; Gómez *et al.*, 2008).

A pesar de la importancia y utilidad del coeficiente de consanguinidad individual, presenta un importante inconveniente que limita su empleo directo en estos fines, particularmente en el que nos ocupa en este texto, la descripción de la estructura genética de poblaciones animales. El problema es la enorme dependencia del parámetro de la profundidad de pedigrí, de manera que un mismo coeficiente puede ser considerado elevado o reducido dependiendo de si la información genealógica es respectivamente escasa o histórica. La transformación de estos coeficientes en el parámetro conocido como tamaño o censo efectivo, nos proporciona una forma de resolver este problema. A continuación se explicará con detalle este parámetro.

3.4.2. El proceso dispersivo

En una población “grande” en ausencia de migración, mutación y selección las frecuencias alélicas permanecen prácticamente inalterables de generación en

generación. Sin embargo en poblaciones “pequeñas” aunque estos tres supuestos se cumplan existen fluctuaciones en las frecuencias alélicas de forma aleatoria debido al muestreo gamético. Estos pequeños cambios en las frecuencias es lo que se denomina *proceso dispersivo* (Kimura, 1983). Para medir el proceso dispersivo comenzaremos definiendo una situación en la que éste se dé en su forma más simple. De este modo, se define la *Población Ideal* (Fisher, 1930; Wright, 1931) como aquella en la que se cumplen las siguientes condiciones: el tamaño de la población base es infinito, se encuentra subdivida en un gran número de subpoblaciones con un número constante de individuos reproductores por subpoblación y generación, el apareamiento es al azar (incluyendo la autofecundación), las generaciones no se solapan, la migración está excluida de forma que las “líneas” se encuentran aisladas, no hay selección en ninguna etapa y no se considera la mutación.

Bajo las condiciones de la población ideal todos los individuos reproductores contribuyen igualmente al conjunto de gametos a partir del cual se formarán los cigotos. La unión de gametos es aleatoria y proporciona los cigotos que darán lugar a los individuos de la siguiente generación. La supervivencia de los cigotos es al azar y, por tanto, la contribución de los padres a la siguiente generación no es uniforme. Puesto que el censo de la población es constante de generación en generación, el número promedio de hijos que alcanza la edad reproductiva es uno por cada padre. De esta forma el esquema de muestreo está representado como un único suceso: la reducción de un gran número de gametos a un número pequeño de descendientes reproductores. No obstante, esta reducción puede tener lugar en varias etapas, y esto no afecta a las consecuencias teóricas que se deducen del número final de descendencia reproductora, siempre y cuando el muestreo en cada etapa sea aleatorio (Falconer y Mackay, 1996).

El proceso dispersivo puede ser concebido y descrito desde dos ópticas distintas:

- A. Como un proceso de muestreo y describirlo en términos de varianza muestral.
- B. Como un proceso de consanguinidad y describirlo en términos de los cambios genotípicos que resultan de los apareamientos.

La primera forma es la más simple e intuitiva ya que se trata de la descripción del propio funcionamiento del proceso, pero la segunda forma nos permite cuantificar de forma más adecuada las consecuencias de dicho proceso (Falconer y Mackay, 1996).

Realizaremos a continuación una descripción del proceso dispersivo desde las dos ópticas centrándonos sobre todo en el proceso de consanguinidad que es el que nos lleva a realizar esta Tesis.

3.4.2.1. El proceso dispersivo desde una óptica de muestreo de alelos

Puesto que se han excluido todos los procesos sistemáticos (migración, mutación y selección) el promedio de frecuencias alélicas para todas las subpoblaciones debe ser igual al inicial.

El cambio de frecuencia alélica resultante del muestreo es un proceso aleatorio, pero su magnitud, con la ayuda de la distribución binomial a la que se adapta, puede predecirse en términos de varianzas de dicho cambio del siguiente modo:

$$\sigma_{\Delta q}^2 = \frac{q(1-q)}{2N} \quad [7]$$

donde q es la frecuencia alélica de un gen de la población base y N el número de individuos reproductores por línea.

3.4.2.2. El proceso dispersivo desde una óptica de consanguinidad

En una población ideal de tamaño N , en la generación inicial existen $2N$ alelos distintos, de manera que la probabilidad de que se produzca homocigosis en la formación de un cigoto es $\frac{1}{2N}$. Este proceso se repite en cada generación, pero

habiéndose de tener en cuenta lo sucedido en generaciones previas, de forma que en la generación t , el coeficiente de consanguinidad puede ser expresado del siguiente modo:

$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1} \quad [8]$$

donde el primer término representa el incremento de consanguinidad producido por la identidad por descendencia de dos copias de un gen que proceden de un individuo de la generación $t-1$ (o endogamia de nueva creación) y el segundo término representa la consanguinidad producida hasta la generación anterior y que se mantiene en los $1 - \frac{1}{2N}$ cigotos restantes en los que no aparece nueva endogamia (endogamia antigua).

De este modo, se define la tasa de endogamia o incremento de consanguinidad relativo a una generación para alcanzar la consanguinidad F_t en la generación t en la población ideal como:

$$\Delta F = \frac{1}{2N} \quad [9]$$

Así, la ecuación [8] se puede expresar también del siguiente modo:

$$F_t = \Delta F + (1 - \Delta F) F_{t-1} \quad [10]$$

Y, despejando el incremento en consanguinidad en esta expresión se obtendría:

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}} \quad [11]$$

Si analizamos esta última expresión, vemos que el ΔF mide la *tasa de consanguinidad* como incremento proporcional, es decir mide el incremento de consanguinidad que se da en una generación t ($F_t - F_{t-1}$ en el numerador), relativo a la consanguinidad que ya había en la generación precedente $t-1$ ($1 - F_{t-1}$ en el denominador).

Se puede operar de forma recursiva la ecuación [11] hasta la generación cero en la que la consanguinidad es nula, lo que llevaría a la siguiente expresión:

$$F_t = 1 - (1 - \Delta F)^t \quad [12]$$

Obsérvese que conociendo el censo de la población ideal se conoce también la tasa de endogamia, de manera que esta expresión nos permite conocer la endogamia que tendría la población en cualquier generación t .

3.4.3. Definición de tamaño efectivo

Es obvio que las condiciones especificadas para la población ideal no se cumplen en las poblaciones reales, y por tanto, el número de individuos reproductores presentes en la población no describe de forma adecuada los efectos de la consanguinidad y de la deriva genética en la mayoría de los casos (Caballero, 1994). De esta manera, el concepto de *tamaño o censo efectivo de población* (N_e) descrito por Wright (1931, 1938, 1939) es la mejor forma de tratar cualquier desviación concreta de la estructura reproductiva ideal. El censo o tamaño efectivo de población se define como el número de individuos que daría lugar a la tasa de

varianza de muestreo ($N_e = \frac{q(1-q)}{2\sigma_{\Delta q}^2}$) o la tasa de consanguinidad ($N_e = \frac{1}{2\Delta F}$)

observados, si dichos individuos se reprodujeran en la manera indicada en la población ideal.

Centrándonos nuevamente en el censo efectivo desde el punto de vista de tasa de consanguinidad, en la población ideal ΔF se relaciona con el censo poblacional N (ecuación 9), por lo que de igual modo el tamaño efectivo de cualquier población se relacionaría con ΔF :

$$\Delta F = \frac{1}{2N_e} \quad [13]$$

N_e puede ser así utilizado para conocer la tasa de consanguinidad y, con la ayuda de la ecuación [12] predecir la consanguinidad en cualquier generación. Obsérvese que la relación inversa también es cierta:

$$N_e = \frac{1}{2\Delta F} \quad [14]$$

A partir de esta expresión, si es posible obtener la tasa de endogamia en una población, se dispone automáticamente del censo efectivo, parámetro de gran interés en la descripción de la estructura genética de poblaciones.

3.4.4. La deriva genética y la utilidad del tamaño efectivo

El proceso dispersivo está descrito en un escenario de varias líneas procedentes de una misma población y tiene cuatro consecuencias (Falconer y Mackay, 1996):

- Una de ellas es la deriva genética que, consiste en el cambio aleatorio de las frecuencias alélicas cuya dirección no se puede predecir pero sí su magnitud.
- La diferenciación entre líneas: los individuos de distintas líneas son cada vez más diferentes, ya que la deriva genética ocurre de forma independiente en cada una de ellas.
- La homogeneización dentro de líneas: la variación genética se reduce progresivamente dentro de cada línea, cuyos individuos se hacen cada vez más parecidos.
- El incremento de la homocigosis: los individuos homocigotos incrementan en detrimento de los heterocigotos.

La deriva genética puede ser medida, al menos en poblaciones no estructuradas, mediante el parámetro censo o tamaño efectivo (Caballero, 1994).

La pérdida de variabilidad genética por generación es inversamente proporcional al tamaño efectivo de la población (Wright, 1931, 1951) y se denomina *tasa de dispersión* en este contexto, aunque como vimos anteriormente se llama también *tasa de endogamia*.

Además de la deriva genética existen otros procesos que conducen a la modificación de la variabilidad genética, como son, la migración, la mutación y la selección.

Como acabamos de ver, en una población cerrada de tamaño finito, en ausencia de mutación, se produce inevitablemente la pérdida de variabilidad genética por generación como consecuencia de la deriva, siendo esta pérdida tanto más importante cuanto menor sea el censo de la población. La migración es en este caso, el único factor capaz de contrarrestar esta pérdida de variabilidad genética. Afortunadamente tasas muy pequeñas de migración son suficientes para contrarrestar las pérdidas de variabilidad genética por deriva (Falconer y Mackay, 1996). Se ha descrito que el hecho de mover un solo individuo de cada subpoblación en cada generación a otra subpoblación al azar (lo que se conoce como modelo isla) es suficiente para mantener niveles aceptables de consanguinidad en cada grupo, a la vez que se mantiene una diferenciación adecuada entre los mismos, la regla de “un migrante por generación” (OMPG) (Wang, 2004).

Por otro lado, en ausencia de migración, si la población mantiene un tamaño constante, y la presión de selección también es constante, se alcanzará un determinado nivel de varianza genética en una situación de pseudoequilibrio, donde la pérdida de varianza genética por selección y deriva se contrarresta con la incorporación de nueva variación por mutación (Lynch, 1996). Si el tamaño efectivo es grande, la deriva genética pierde importancia, y el nivel de varianza genética vendrá determinado por un equilibrio mutación-selección.

El tamaño o censo efectivo poblacional es entonces considerado como el parámetro de referencia en el mundo de la conservación, hasta el punto que sirve de referencia en organismos internacionales como la FAO para decidir qué poblaciones se encuentran en “peligro de extinción”, “en desarrollo” o deben ser consideradas “razas de fomento”. Desafortunadamente el verdadero valor del parámetro es desconocido en poblaciones reales, y diversas estimaciones del mismo pueden conducir a valores dispares en función de la metodología utilizada, basándose las más habituales en el desigual número de descendientes por individuo y en los incrementos de consanguinidad medio de los períodos más recientes.

3.4.5. Estimación del tamaño efectivo

3.4.5.1. Incumplimiento de alguna de las condiciones de la población ideal

Dado que las condiciones de la población ideal son muy exigentes y es prácticamente imposible que se cumplan en poblaciones reales, se ha hecho un esfuerzo en aproximar el tamaño efectivo en diversas situaciones. Revisiones bastante completas pueden encontrarse en Caballero (1994) y Falconer y Mackay (1996), aunque extractaremos a continuación algunas de las más utilizadas.

3.4.5.1.1. Exclusión de la autofecundación

Es la excepción más simple de la población ideal. Dado que en poblaciones animales se trabaja habitualmente con organismos bisexuales, y en éstos el proceso de autofecundación es imposible, resolver esta excepción podría ser de utilidad. En este tipo de poblaciones, la probabilidad de que dos gametos que se unan para producir un cigoto en la generación t lleven copias de un gen de un

individuo en la generación $t-2$ es $\frac{1+F_{t-2}}{2N}$ y la probabilidad de que lleve copias

idénticas de un gen procedente de diferentes individuos es $\left(1 - \frac{1}{N}\right)F_{t-1}$. Por

tanto, la consanguinidad en la generación t sería expresada por la siguiente ecuación:

$$F_t = \frac{1+F_{t-2}}{2N} + \left(1 - \frac{1}{N}\right)F_{t-1} \quad [15]$$

Bajo esta situación, la deriva genética comienza dos generaciones antes que la consanguinidad. El incremento de consanguinidad vendría dado por $\Delta F \approx \frac{1}{2N+1}$ (Wright, 1969) y despejando el tamaño efectivo sería $N_e \approx N + \frac{1}{2}$.

Esta aproximación introduce un error muy pequeño en el cálculo de ΔF , a menos que N sea muy pequeño.

3.4.5.1.2. Exclusión del apareamiento fraternal

Esta situación, en la que se evitan cruces entre individuos relacionados, se trata de manera similar a la anterior y se llega a las siguientes expresiones para la tasa de endogamia y el censo efectivo: $\Delta F \approx \frac{1}{2N+4}$ y $N_e \approx N + 2$.

3.4.5.1.3. Número diferente de padres y madres

Es muy frecuente que el número de individuos de los dos sexos no esté equilibrado. En la mayoría de las poblaciones domésticas es más económico usar menos padres que madres. Sin embargo, cualquiera que sean sus números relativos, los dos sexos contribuyen con la misma proporción de genes a la siguiente generación. La probabilidad de que dos genes en $t-1$ que se unan para producir un individuo en la generación t , procedan ambos, de machos de la generación $t-2$ es $\frac{1}{4}$, y la probabilidad de que procedan del mismo macho sería $\frac{1}{4N_m}$. Lo mismo ocurriría para las hembras, de manera que la probabilidad de

que dos genes que se unen para producir un cigoto en la generación t , procedan de un mismo individuo de la generación $t-2$, independientemente de su sexo es $\frac{1}{4N_m} + \frac{1}{4N_h}$ y por tanto:

$$\frac{1}{N_e} \approx \frac{1}{4N_m} + \frac{1}{4N_h} \rightarrow N_e \approx \frac{4N_m N_h}{N_m + N_h} \rightarrow \Delta F \approx \frac{1}{8N_m} + \frac{1}{8N_h} \quad [16]$$

De nuevo, esta aproximación tendría un error pequeño salvo que, el número de machos y hembras sea muy pequeño. Además, se pone de manifiesto que el tamaño del sexo menos numeroso es el que va a tener mayor impacto en la tasa de endogamia. Si en una población los machos y hembras se mantuvieran siempre en la misma proporción, el N_e sería máximo e igual a N. Sin embargo, si la población estuviera compuesta por un número indefinido de hembras pero con un solo macho el tamaño efectivo se aproximaría a cuatro (Falconer y Mackay, 1996).

3.4.5.1.4. Número diferente de reproductores en generaciones sucesivas

Una de las condiciones de la población ideal es que el número de individuos reproductores es constante a través de todas las generaciones. Si la tasa de

consanguinidad, en cualquier generación, viene dada por la expresión [12] y el número de progenitores no es constante, la tasa media de consanguinidad es el valor medio de $\frac{1}{2N}$ en las generaciones sucesivas, y el censo efectivo es la media armónica del número de progenitores en cada generación. Por tanto, en un periodo de t generaciones (Wright, 1938; Crow y Kimura, 1970) tendríamos:

$$\frac{1}{N_e} \approx \frac{1}{t} \sum_{i=1}^t \frac{1}{N_i} \quad [17]$$

Esta expresión puede ser extendida para ambos sexos del siguiente modo (Chia y Pollak, 1974):

$$\frac{1}{N_e} \approx \frac{1}{t} \sum_{i=1}^t \left(\frac{1}{4N_{m,i}} + \frac{1}{4N_{h,i}} \right) \quad [18]$$

Hay que hacer dos consideraciones, la primera es que el máximo valor de censo efectivo se alcanzaría cuando el tamaño de la población se mantuviera constante con el paso de las generaciones. Y segunda, que el censo efectivo está muy afectado por los “cuellos de botella”, que causan un gran incremento en la consanguinidad que *a posteriori* no es recuperado a pesar de que se aumente el tamaño poblacional.

3.4.5.1.5. Distribución no aleatoria del número de hijos por familia

En la población ideal todos los padres tienen la misma probabilidad de dejar descendientes que se usarán como reproductores en la generación siguiente, de forma que las diferencias entre sus contribuciones son debidas al azar. En esta población ideal cada pareja proporciona para la siguiente generación dos hijos, un macho y una hembra, pero con variación aleatoria del número de hijos por familia, lo que se correspondería con una distribución binomial. Si la población no es muy pequeña, la distribución binomial es muy parecida a la distribución de Poisson, donde la varianza es igual a la media, es decir, $V_k = k = 2$, cuando los padres tienen igual probabilidad de contribuir a la siguiente generación, siendo k el número de descendientes por pareja y V_k la varianza del tamaño familiar.

Sin embargo, en las poblaciones reales esto no ocurre así, ya sea por diferencias en la fertilidad, en la supervivencia de la descendencia o simplemente por imposición por parte de los criadores (Caballero, 1994). Esta variación entre las contribuciones de los padres, produce una mayor variación del número de hijos por familia, y eso tiene como consecuencia que una mayor proporción de individuos en la generación siguiente provenga de un número reducido de padres, y de esta forma, el censo efectivo se reduce y la varianza del número de hijos por familia es mayor que dos. La forma en que la varianza influye sobre el tamaño efectivo puede deducirse considerando la probabilidad de que un cigoto sea homocigoto idéntico por descendencia, de manera similar a como se dedujo anteriormente el incremento en consanguinidad, y la relación es:

$$N_e \approx \frac{4N}{V_k + 2} \quad [19]$$

La expresión anterior debe ser modificada si los apareamientos no son monógamos. Si los machos pueden aparearse con más de una hembra, la varianza probablemente será diferente para machos y hembras y usaremos la ecuación:

$$N_e \approx \frac{8N}{V_{km} + V_{kh} + 4} \quad [20]$$

donde V_{km} y V_{kh} son las varianzas del número de hijos machos y hembras respectivamente (Hill, 1979).

En ocasiones interesa realizar estrategias de cruzamiento para producir, por ejemplo, un mínimo incremento de consanguinidad. En estos casos la varianza del tamaño familiar se modifica intencionadamente escogiendo el mismo número de individuos de todas las familias, de manera que no haya variación en el número de hijos por familia ($V_k=0$). Utilizando la ecuación [19], y teniendo en cuenta la relación exacta quedaría:

$$N_e = 2N - 1 \quad [21]$$

Si además hubiera diferencias en el número de individuos de cada sexo, la varianza de hijos por familia podría hacerse cero escogiendo como futuros padres un macho de la descendencia de cada padre y una hembra de la descendencia de cada madre. La tasa de consanguinidad se escribiría entonces (Gowe *et al.*, 1959):

$$\Delta F = \frac{3}{32N_m} + \frac{1}{32N_h} \quad [22]$$

donde N_m y N_h son los números reales de padres y madres, respectivamente, donde las hembras son más numerosas que los machos.

3.4.5.1.6. Solapamiento generacional

El solapamiento generacional es un hecho habitual en las poblaciones de animales domésticos, es decir, que haya individuos de diferentes edades y que sus descendientes puedan actuar como reproductores en diferentes generaciones y no sólo en la siguiente a la suya, siendo imposible distinguir generaciones discretas en la población. Además existen diferencias en la longevidad, por lo que no todos los individuos dejarán descendientes durante la misma cantidad de tiempo, ya que los más longevos tendrán más oportunidades de dejar descendencia. Esto afecta nuevamente a la varianza del tamaño familiar y al censo efectivo pudiéndose utilizar modificaciones de las expresiones [19] y [20]. La novedad aquí radica en la necesidad de aproximar el número total de individuos reproductores por generación. Este número puede calcularse contando el número de reproductores en un periodo de tiempo de referencia (N_c , en, por ejemplo un

año), y multiplicando después por el intervalo generacional medido en periodos de referencia (l , por ejemplo en años) (Falconer y Mackay, 1996). La expresión [19] quedaría entonces:

$$N_e \approx \frac{4N_c l}{V_k + 2} \quad [23]$$

El mismo razonamiento podría hacerse para el caso de diferencias en el número de machos y la varianza del número de descendientes, donde habría que tener en cuenta las varianzas de hijos e hijas para cada progenitor (macho y hembra), así como las covarianzas entre ellos, lo que llevaría a transformar la expresión [20] a:

$$\begin{aligned} \frac{1}{N_e} = & \frac{1}{16ML} \left[2 + \sigma_{mm}^2 + 2 \left(\frac{M}{F} \right) \text{cov}(mm, mf) + \left(\frac{M}{F} \right)^2 \sigma_{mf}^2 \right] \\ & + \frac{1}{16FL} \left[2 + \left(\frac{F}{M} \right)^2 \sigma_{fm}^2 + 2 \left(\frac{F}{M} \right) \text{cov}(fm, ff) + \sigma_{ff}^2 \right] \end{aligned} \quad \text{(Hill, 1979) [24]}$$

donde M y F son el número de machos y hembras utilizados posteriormente en reproducción en un determinado periodo, L el intervalo generacional, σ_{mm}^2 y σ_{mh}^2 son las varianzas del número de los descendientes machos y hembras de los padres, σ_{hm}^2 y σ_{hh}^2 son las varianzas del número de los descendientes machos y hembras de las madres, y $\text{cov}(mm, mh)$ y $\text{cov}(hm, hh)$ las respectivas covarianzas.

3.4.5.1.7. Mutación, migración y selección

Además del proceso dispersivo, responsable de la deriva genética, en las poblaciones de animales domésticos existen otros procesos sistemáticos que afectan al tamaño efectivo como son la migración, la mutación y la selección. Los procesos sistemáticos tienden a llevar las frecuencias alélicas a un equilibrio estable, mientras que los procesos dispersivos tienden a alejarlas del valor de equilibrio. La tendencia de los procesos sistemáticos a cambiar la frecuencia alélica hacia su valor de equilibrio se hace más patente conforme la frecuencia se desvía de ese valor. Por esta razón hay un punto de equilibrio en el cual la dispersión de las frecuencias alélicas es contrarrestada por los procesos sistemáticos (Falconer y Mackay, 1996).

Dentro del proceso mutacional, encontramos la mutación neutra no recurrente que se trata de mutaciones puntuales. Cada mutación en un *locus* autosómico tiene un 50% de probabilidad de sobrevivir a la siguiente generación. Unas cuantas perdurarán y se extenderán en la población por medio del proceso dispersivo. En cuanto a la selección natural (la artificial se ha tratado en las expresiones desarrolladas anteriormente) parece razonable que actúe a favor de los heterocigotos, ya que los alelos deletéreos actúan normalmente en homocigosis produciendo individuos no viables (Falconer y Mackay, 1996).

Nos centraremos en este apartado en la mutación y migración recurrentes, al ser donde se establece un desarrollo matemático relacionado con el censo efectivo.

En los casos de la mutación y migración recurrentes, ambos procesos pueden estudiarse conjuntamente, ya que afectan a las frecuencias alélicas del mismo modo. Volviendo a la población subdividida en líneas, cada una con un censo efectivo N_e , definimos m como la proporción de individuos inmigrantes reproductores por generación procedentes de otras líneas al azar con respecto al total de reproductores en la línea. Definimos también u y v como las tasas de mutación en los dos sentidos entre los dos alelos de un *locus*. Cuando la mutación y migración logran equilibrar la pérdida de variabilidad genética causada por el proceso dispersivo, el coeficiente de consanguinidad permanece estable en la población, relacionándose todos estos parámetros de la manera siguiente (Falconer y Mackay, 1996):

$$F \approx \frac{1}{4N_e(u + v + m) + 1} \quad [25]$$

Habitualmente las tasas de mutación recurrente resultan despreciables en relación a la tasa de migración, de manera que si ambos procesos se tratan conjuntamente la expresión anterior quedaría ($u + v = 0$):

$$F \approx \frac{1}{4mN_e + 1} \quad [26]$$

3.4.5.2. Estimación del tamaño efectivo en Poblaciones Reales

En el apartado previo se han ido considerando casos particulares en los que había pocos supuestos de la población ideal que no se cumplían, de manera que ha sido posible encontrar una solución válida desde el punto de vista de la genética de poblaciones teórica. A continuación realizaremos una revisión sobre las metodologías propuestas para abordar la estimación del tamaño efectivo en poblaciones reales de animales domésticos con un archivo genealógico. Estas poblaciones se caracterizan por presentar prácticamente todas las excepciones al modelo de la población ideal en mayor o menor grado.

Todos los métodos que se muestran a continuación se basan en regresión, por ser estas técnicas las más universalmente aceptadas.

Por simplicidad teórica desarrollaremos la metodología para el caso de generaciones discretas y veremos después su uso bajo solapamiento generacional.

3.4.5.2.1. Estimación del tamaño efectivo en poblaciones con generaciones discretas

Dado que en este caso es posible asignar cada individuo a una generación, se puede calcular una regresión lineal de la consanguinidad F sobre la generación t , de modo que el incremento de consanguinidad producido entre la generación t y la $t-1$ vendrá dado por la pendiente de la recta (b). Si la consanguinidad media en la

población es muy baja podemos aproximar el denominador a 1, despreciando así la consanguinidad en la generación previa (F_{t-1} en el denominador):

$$b = F_t - F_{t-1} \approx \Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}} \quad [27]$$

y por tanto, el censo efectivo se podría calcular del siguiente modo:

$$N_e = \frac{1}{2b} = \frac{1}{2\Delta F} \quad [28]$$

Si la consanguinidad en la población no es despreciable, tendríamos que obtener el valor de la consanguinidad en la generación precedente como el valor de la consanguinidad en la población de referencia menos el incremento que se da en una generación: $F_{t-1} = F_t - b$. En este caso el incremento de consanguinidad entre las dos generaciones vendría dado por:

$$\Delta F = \frac{b}{1 - (F_t - b)} \quad [29]$$

y utilizaríamos este valor para calcular el censo efectivo.

El incremento de los coeficientes de consanguinidad medio por generación no es lineal para una política de apareamientos estable, pero habitualmente puede considerarse como tal en los períodos en los que se suelen llevar a cabo este tipo de estudios. Sin embargo, en caso de que no fuese así, el incremento de los coeficientes de consanguinidad se podría obtener mediante regresión logarítmica, ya que la ecuación [12] puede escribirse de este modo:

$$1 - F_t = \left(1 - \frac{1}{2N_e} \right)^t \quad [30]$$

A partir de la regresión logarítmica de $(1 - F_t)$ sobre t , se puede despejar el valor del censo efectivo (N_e).

Nótese también que en el caso de disponer de generaciones discretas se puede calcular directamente el censo efectivo que se ha dado en cada una de las generaciones mediante el empleo directo de la expresión [11] sin necesidad de hacer uso de estimaciones basadas en procedimientos estadísticos.

3.4.5.2.2. Estimación del tamaño efectivo en escenarios de solapamiento generacional

La estimación del tamaño efectivo en situación de solapamiento generacional, situación clásica en la gran mayoría de las poblaciones reales de animales domésticos, presenta la dificultad añadida de no poder asignar a cada individuo una generación concreta. Una primera aproximación sería el empleo de la expresión [24] (Hill, 1979). Desafortunadamente esta expresión únicamente está

basada en las diferencias en tamaño de familia y tiende a sobreestimar el parámetro.

Una alternativa sería el uso del intervalo generacional (l). Inicialmente se calcularía el coeficiente de regresión lineal (b) del coeficiente de consanguinidad individual sobre el año de nacimiento y se corregiría después este incremento por el intervalo generacional en años (Gutiérrez *et al.*, 2003). Obsérvese que este procedimiento puede ser mejorado si las fechas de nacimiento y los intervalos generacionales se manejan en unidades de tiempo más pequeñas. De forma paralela al punto anterior, el incremento de consanguinidad media producido entre la generación t y la $t-1$, si se asume despreciable la consanguinidad en la generación previa, sería:

$$F_t - F_{t-1} = lb \approx \Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}} \quad [31]$$

También de forma paralela, si la consanguinidad en la generación previa no es despreciable, el incremento de consanguinidad sería:

$$\Delta F = \frac{lb}{1 - (F_t - lb)} \quad [32]$$

Se obtendría así, en cada uno de los dos casos, el incremento de consanguinidad y a partir de él, el censo efectivo.

Una vez más, de forma paralela a lo comentado en el apartado anterior, si el incremento en consanguinidad no puede aproximarse linealmente (normalmente por cubrir un elevado número de generaciones), el tamaño efectivo puede estimarse a partir de la regresión logarítmica de $(1 - F)$ sobre el año de nacimiento a (Pérez-Enciso, 1995):

$$1 - F = \left(1 - \frac{1}{2N_e}\right)^a \quad [33]$$

En esta expresión se obtendría un falso tamaño efectivo que sería transformado a otro más correcto mediante su cociente por el intervalo generacional.

Finalmente otra alternativa consistiría en utilizar el número de generaciones equivalentes (Maignel *et al.*, 1996):

$$t_i = \sum_1^a \left(\frac{1}{2}\right)^{n_a} \quad [34]$$

En esta expresión a es el número total de antepasados del animal i , y n_a el número de generaciones que separan al individuo de cada antepasado.

Así, una vez asignado el individuo a una generación no discreta, se puede proceder a la utilización de procedimientos basados en regresión, descritos en el apartado anterior.

3.4.5.3 Estimación del censo efectivo con ausencia de información genealógica

3.4.5.3.1. Utilización de la información de campo. Encuestas a ganaderos

Cuando no se disponen de registros genealógicos y las características de la población lo permiten, se pueden elaborar encuestas en ganaderías recogiendo datos relacionados con el manejo reproductivo de los animales, número de reproductores en uso, número de padres y madres, edad media de desvieje, tasa de reposición, número de hijos por macho, etc. Todos ellos son parámetros demográficos que nos ayudan a determinar el tamaño efectivo de la población. Así se pueden estimar las varianzas y covarianzas de la expresión [24] a partir de encuestas recogidas durante una estación de cría en una determinada población y estimar el censo efectivo sin necesidad de herramientas genealógicas (Ron Garrido *et al.*, 2006).

3.4.5.3.2. Utilización de la información molecular

En los casos en los que la cantidad y calidad de pedigrí que se dispone de una población es escasa, el uso de herramientas genealógicas tiende a sobreestimar el parámetro. En estas ocasiones, y en ausencia total de pedigrí, se pueden utilizar herramientas moleculares para la determinación del tamaño efectivo. A pesar de que estos estudios son escasos en las poblaciones de animales domésticos, existen metodologías basadas en el desequilibrio de ligamiento (Hill, 1981), en la variación de las frecuencias alélicas en un determinado periodo separado por un determinado número de generaciones (Waples, 1989) y en análisis bayesianos basados en la teoría de la coalescencia (Berthier *et al.*, 2002).

ARTÍCULOS



4.1. ARTÍCULO 1: POPULATION HISTORY AND GENETIC VARIABILITY IN THE SPANISH ARAB HORSE ASSESSED VIA PEDIGREE ANALYSIS

Livestock Science (2008), 113: 24-33

I. Cervantes^{a,*}, A. Molina^a, F. Goyache^b, J.P. Gutiérrez^c, M.Valera^d

^aDepartment of Genetics. University of Córdoba. Ctra. Madrid-Cádiz, km 396a, 14071 Córdoba. Spain.

^bSERIDA-Somío, C/Camino de los Claveles 604, E-33203 Gijón. Spain.

^cDepartment of Animal Production. Faculty of Veterinary. UCM. Avda. Puerta del Hierro s/n, 28040 Madrid. Spain.

^dDepartment of Agro-Forestral Sciences. EUITA. University of Sevilla. Ctra. Utrera km 1, 41013 Sevilla. Spain.

Received 4 september 2006; received in revised form 26 January 2007; accepted 14 February 2007

* Corresponding author: I. Cervantes. Department of Genetics, University of Córdoba, Edif. Gregor Mendel P. Baja, Ctra. Madrid-Córdoba km. 396^a, E-14071, Córdoba, Spain. Tel.: +34 957218735; fax: +34 957218707
E-mail address: v82cenai@gmail.com

Abstract

The studbook of the Spanish Arab Horse was analysed in order to assess the available genetic variability in the population via genealogical analyses and to ascertain the influence of different countries in the formation of the population. The Spanish Arab horse studbook could be traced back to the late 19th century with a total of 18,880 animals. A 15.7 % (2,965) of the registered individuals were Arab horses imported from foreign populations. The total number of founders was 1,626: from these, 1,495 individuals were identified as imported from foreign countries. The sum of the genetic contributions of the founders to the population revealed that those imported from the 'Middle East', Poland, the United Kingdom and Egypt had a higher importance in the formation of the Spanish Arab Horse, contributing, respectively, 36.4%, 18.7%, 11.9% and 9.1% of the genetic variability. The effective number of founders was 38.6 and the effective number of ancestors was 19 thus characterising the abusive use of a little number of individuals for reproduction. This was due to a breeding policy that included mating between relatives, but also due to the limited availability of reproductive individuals after the Spanish Civil War. The average inbreeding of the population was 7.0% (9.8% for the individuals born during 1995-2004). Individuals with very high levels of inbreeding ($\geq 12.5\%$) were 17.7% in the whole population and 26.7% of the individuals born in the last decade. Positive genealogical F_{IS} near 2% characterised a breeding policy including matings between relatives. No clear within-population differentiation (inferred via genealogical F_{ST}) could be assessed between sire lines ($F_{ST} = 0.01$) and no differentiation between geographical areas existed. Overall, the Spanish Arab Horse can be described as a population without genetic subdivision and with a mating policy and historical constraints that reduced the available genetic variability regardless of the relatively large size of its founder population. The obtained information will be useful to implement a selection program in the Spanish Arab Horse.

Keywords: Spanish Arab horse; genetic variability; pedigree analysis; population history; within-population differentiation

1. Introduction

The Arab breed is one of the most influential horse breed in the World (Bowling and Ruvinsky, 2000a). It is distributed worldwide and has been involved in the formation of many other horse breeds, such as the Thoroughbred (Bowling and Ruvinsky, 2000b) or the Lipizzan (Zechner *et al.*, 2002). At present, the Arab horse breed has the third census size in Spain, totalling 8.1% of the horses in our country and when the Arab and the Arab crossbreed horse breeds (Hispano-Arab and Anglo-Arab) are considered it moves up to the second place among horse populations in Spain after the Andalusian (Spanish Purebred) horse (MAPyA, 2003). The Spanish Arab studbook was founded in 1847 by the Spanish Ministry of War using, mainly, individuals imported from the Middle East and from various European countries (Maxwell, 1995). During the first third of the 20th century a number of private studs were formed in Southern and Central Spain from individuals purchased in the Military Stud of Jerez de la Frontera and foreign countries such as Poland, the United Kingdom, France, Egypt and the area of the Arabian Peninsula (Fenau, 1995; Maxwell, 1995). Since the 70's a significant number of Spanish Arab Horse studs have been founded in other Spanish regions such as Catalonia, Balearic Islands and the Cantabrian provinces. The Spanish Arab studbook remains open to include Arab horses from foreign countries that are not always registered with full genealogies in the Spanish Arab Studbook. As

happened in other Spanish horse breeds (Valera *et al.*, 2005) the interest in the Spanish Arab Horse has increased during the past decades with the number of registrations in the studbook. Currently, the Spanish Arab Horse Breeder Association (AECCA) is developing a selection scheme with the aim of improve conformation traits and horse performance, mainly, in endurance rides.

The assessment of the within-population genetic variability, population structure and gene flow is necessary during the implementation of selection programs to establish an appropriate management of the genetic stock in order to enlarge genetic basis for selection. In this respect genealogical tools are important to describe these subjects (Moureaux *et al.*, 1996; Głażewska and Jezierski, 2004). Moreover, demographical analyses may also help to understand important circumstances affecting the genetic history of the population (Zechner *et al.*, 2002; Valera *et al.*, 2005).

The aim of this work is to analyse the pedigree information of the Spanish Arab Horse included in its studbook in order to quantify the available genetic variability in the population in terms of inbreeding levels and founder contributions to the present population. The study will also focus on analysing the influence of animals from foreign countries in the Spanish Arab Horse and to assess a possible population structure. The implications on the genetic history of the Spanish Arab Horse population over a century of breeding will also be discussed.

2. Material and methods

Information registered in the studbook of the Spanish Arab Horse from its foundation in 1847 to December 2004 has been analysed. Data included 18,880 (8,893 males and 9,987 females) registered animals. Information from current and older stud books of major Spanish Arab studs were primary sources of information for the identification of the origin of the individuals imported from foreign countries (FC). Important secondary source were the books of Fenaux (1995), Maxwell (1995) and Steen (2006). Altogether, FC individuals included those animals actually imported, their ancestors included in the Spanish Arab Horse studbook regardless of location in the Spanish territory and those individuals born in Spain but registered as imported 'in uterus'. Those horses from which was known that were bred in Spain before their registration in the studbook and those individuals born in Spain from registered individuals regardless of the country of origin of their parents were classified as Spanish native individuals (SN). Following Fenaux (1995), those horses assigned to the old 'Duque de Veragua' stud, from which genealogical information was lost during the Spanish Civil War, were also considered as SN individuals. Following Moureaux *et al.* (1996) ancestors with both parents unknown in the available database (managed and provided by the Spanish Ministry of Defence) were considered founders. If one parent is known, the unknown parent was considered a founder (Lacy, 1989). Notice that, using this criterion, some FC individuals can be considered founders for the present study regardless their parents are registered in foreign Arab horse studbooks.

Traditionally, a few familiar lines were identified in the Spanish Arab Horse according to the historical importance of a given stallion in the formation of the

population (Fenaux, 1995): *Ursus*, *Seanderich*, *Razada* and *Bagdad*. These sires have been traced forward in the pedigree in order to identify 'sire lines' defined as unbroken descent through male animals only from an ancestor to a descendant (Cunningham *et al.* 2001; Royo *et al.*, 2007).

The analyses listed below were carried out for the whole population. However some informative parameters were also computed for a group of animals formed by the animals born between 1995 and 2004, in order to approach the last generation. Throughout the text this group of individuals will be named as 'reference population'.

2.1. Demographic parameters

The pedigree completeness level was assessed by computing for the whole pedigree as the proportion of ancestors known per generation for each offspring (McCluer *et al.*, 1983). Also, the number of equivalent generations (g_e) in the pedigree was computed as the sum of $(1/2)^n$ where n is the number of generations separating the individual to each known ancestor (Boichard *et al.*, 1997).

2.2. Probability of gene origin

The genetic representation of the founder population in the Spanish Arab Horse was assessed using the following parameters:

- Effective number of founders (f_e). This parameter is the reciprocal of the probability that two genes drawn at random in the studied population originate from the same founder (James, 1972) and is computed from the genetic contribution of founders to the descendent gene pool of the population (Lacy, 1989).
- Effective number of ancestors (f_a). To compute this parameter, the ancestors explaining a percentage of population higher than their parents were identified, and only their marginal contribution, that one not explained by other ancestors previously chosen, was considered. This parameter complements the information offered by the effective number of founders accounting for the losses of genetic variability produced for the unbalanced use of reproductive individuals producing bottlenecks. It could be defined as the minimum number of ancestors, not necessarily founders, explaining the complete genetic diversity of a population (Boichard *et al.*, 1997). The genetic contributions of the ancestors has been averaged by decade of birth of the ancestors in order to identify populations bottlenecks.

2.3. Inbreeding and Average Relatedness Coefficient

The level of heterozygosity of the population was assessed using the following parameters:

- Individual inbreeding coefficient (F) defined as the probability that an individual has two genes identical by descent (Malécot, 1948).
- Average Relatedness coefficient (AR) of each individual defined as the probability that an allele randomly chosen from the whole population belongs to a given

animal (Goyache *et al.*, 2003; Gutiérrez *et al.*, 2003). The AR Coefficient for each individual in the pedigree is computed as the average of the coefficients in the row correspondent to the individual in the numerator relationship matrix **A** (Gutiérrez *et al.*, 2003). Notice that both alleles from the individual are included in the population. AR is twice the mean coancestry between a given animal and all animals in the population including itself. Thus the AR accounts simultaneously for the coancestry and inbreeding coefficient. Notice that the AR coefficient of a founder is its genetic contribution to the entire pedigree.

In order to ascertain the genetic contribution (Lacy, 1989) over the time of the historically sire lines, we traced forward the genetic representation of *Ursus*, *Seanderich*, *Razada* and *Bagdad* in all the individuals of the population and averaged them by decade of birth of their descendants.

2.4. Genetic structure

Genetic structure of the Spanish Arab Horse population was assessed using F-statistics (F_{IT} , F_{ST} , F_{IS} ; Wright, 1978). F_{IT} is the inbreeding coefficient of the individual relative to the whole population, F_{ST} is the average inbreeding of the sub-population relative to the whole population and F_{IS} is the inbreeding coefficient of the individual relative to its own subpopulation (Falconer and Mackay, 1996). These parameters were computed following Caballero and Toro (2000, 2002) as

$$F_{IT} = \frac{\bar{f}_0 - \bar{f}}{1 - \bar{f}}, F_{ST} = \frac{\bar{f}_0 - \bar{f}}{1 - \bar{f}} \text{ and } F_{IS} = \frac{\bar{f}_0 - \bar{f}}{1 - \bar{f}} \text{ where } \bar{f}_0 \text{ and } \bar{f} \text{ are, respectively,}$$

the mean coancestry and the inbreeding coefficient for the entire population, and, \bar{f} the average coancestry for the subpopulation so that $(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST})$. This analysis was carried out considering two different levels of population subdivision: the sire line level and the geographical area level. At the sire line level the analysis included the traditional *Ursus*, *Seanderich*, *Razada*, and *Bagdad* lines besides the 'Unsigned' stallions as subpopulations. At the geographical area level the analysis was carried out considering as independent populations: a) the Spanish Mediterranean Levant and Balearic Islands; b) Northern Spain; c) Andalusia; d) Centre of Spain; and e) the Military Stud which was treated as an independent population because of its breeding importance.

Most parameters have been computed using the program ENDOG v 3.0 (Gutiérrez and Goyache, 2005) and other programs developed by the authors.

3. Results

3.1. Demographic Analyses

Table 1 gives the frequency of individuals registered in the Spanish Arab studbook per year of birth (grouped in decades) of the animals. Thirty three percent of the individuals were registered during the last decade (1995-2004), illustrating a demographic evolution caused by an increasing interest of the breeders jointly with the improvement of recording activity. The number of registrations decreased during the decade 1935-44 due to the consequences of the Spanish Civil War

(1936-1939). Among the registered individuals, the proportion of males and females used for reproduction was 12.6% and 27.9%, respectively.

Considering the progeny size, most stallions (71.2%) had an offspring size between 1 and 5. However, the use of the males for reproduction is not balanced and a little number of sires (27; 1.1% of the stallions) had more than 70 descendants.

The total number of FC (Foreign Countries) individuals registered in the Spanish Arab studbook was 2,965 (15.7 % of the total). The contribution of the FC individuals to the studbook is high; a 47.9% of the SN (Spanish Native) individuals registered during the period 1905-1934 had FC parents. During the period 1955-2004 this proportion was of 16.4% (Table 1), in spite the fact that most of FC animals (74.5%) were imported between 1945 and 1994. The total number of founders was 1,626, 1,495 of them being FC founders. The founders of Spanish origin totalled 131 individuals. The country of origin of 333 FC founder individuals could not be identified with certainty. The other founder FC individuals (1,162) were imported from a total of 29 countries. The percentage of founder individuals registered in the studbook according to their country of origin is shown in Figure 1 (see Plot A). The United Kingdom was the country with the highest number of FC founders individuals (12.5%) followed by Poland (10.1%). The other countries that significantly contributed with founder FC individuals to the Spanish Arab studbook were: USA, Egypt, "Middle East" (which is the area including the Arabian Peninsula, and the present Jordan, Syria and Iraq) and France, supplying, respectively, 9.3%, 8.5%, 6.9% and 5% of the FC founders individuals. The SN founders were 8.1% of the total.

The quality of the available pedigree information is illustrated in Figure 2. Completeness level for the whole pedigree was of 92.0%, 86.6% and 80.8 % for, respectively, the first, second and third known generations. These levels drop to values below 40% after the eighth generation. When only the animals born in the reference population are considered the completeness of the pedigree information was higher than 90% until the sixth parental generation. In any case no more than 8 full generations (both parents known) could be traced for any individual in the pedigree. The situation above is also reflected by the number of equivalent generations; this parameter was 7.9 for the individuals born and registered in the last decade whilst for the whole pedigree reached values of 5.7.

Table 1. Description of the Spanish Arab Horse population by decade of birth of the individuals registered in the studbook.

Decade	Total Registrations	Number individuals with foreign parents (%)	Genetic Contribution ^a	AR ^b	F ^c	F ₅ ^d
<1895	86	86 (100)	0.3	0.2	0	0
1895-1904	172	25 (14.5)	12.5	0.4	0	0
1905-1914	158	105 (66.5)	7.6	1.9	0.1	0.1
1915-1924	434	262 (60.4)	5.1	3.0	1.3	1.3
1925-1934	791	296 (37.4)	38.2	4.4	2.2	2.1
1935-1944	615	65 (10.6)	22.1	4.8	2.2	2.1
1945-1954	1014	84 (8.3)	3.3	4.7	2.5	2.4
1955-1964	1095	159 (14.5)	2.9	5.3	3.2	2.9
1965-1974	1313	370 (28.2)	3.3	6.2	3.7	3.1
1975-1984	2766	635 (23.0)	3.2	9.2	6.4	4.9
1985-1994	4196	595 (14.2)	1.1	11.2	9.3	6.1
1995-2004	6240	805 (12.9)	0.4	11.6	9.8	5.2

^aAverage genetic contribution of the Boichard *et al.*'s (1997) ancestors (%), ^bmean average relatedness values (%), ^caverage inbreeding computed using the total pedigree (%), ^daverage inbreeding computed using only the last 5 generations (%).

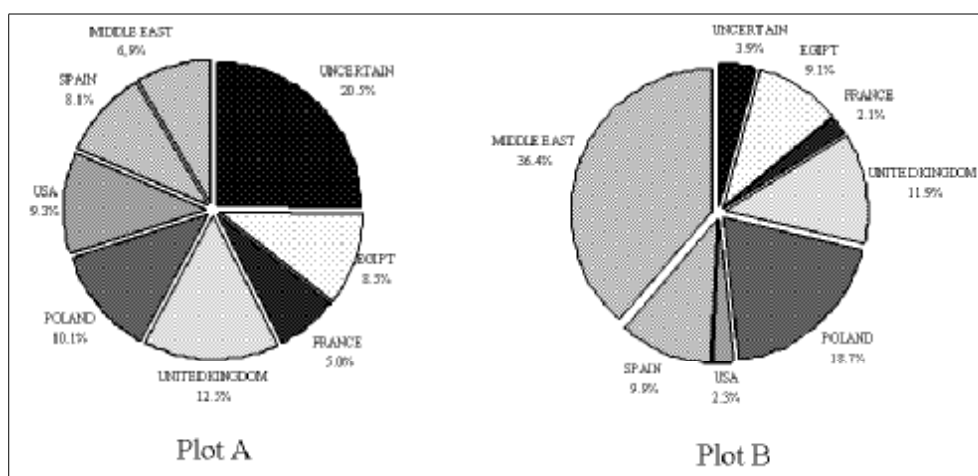


Figure 1. Contributions (in percentage) of different countries (or areas) of origin to the Spanish Arab studbook. Plot A shows the percentage of founder individuals registered in the studbook whilst Plot B gives the sum of the contributions to the genetic variability of the population of these founder individuals. The founder individuals with no clear origin have been classified as 'Uncertain'.

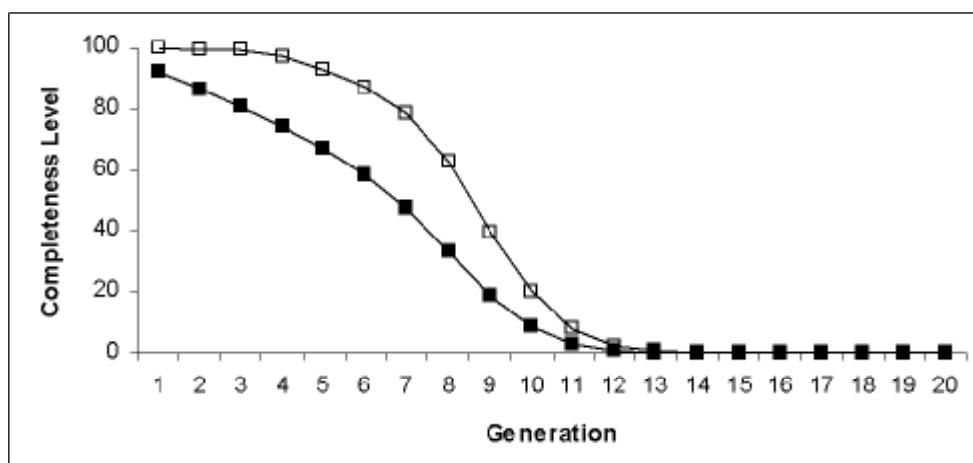


Figure 2. Average percentage of ancestors known per generation, (generation 1 = parents, generation 2 = grandparents, etc.) in Spanish Arab Horse. Information is provided for the whole population (in black squares) and for the last decade (open squares).

3.2. Probability of gene origin

Parameters characterising the genetic variability of the population in terms of gene origin for the whole and the reference populations are given in Table 2. A total of 1,626 (1,385 with both parents unknown and 241 with one parent unknown) individuals were identified as founders. The effective number of founders for the whole population (38.6) was particularly small accounting for a 2.4% of the founders.

Table 2. Parameters characterising the probability of gene origin in the Spanish Arab Horse population for the whole pedigree and for the reference population (animals born in the last decade).

	Whole population	Reference population
Total number of animals in the population	18,880	6,240
Animals with both parents unknown	1,385	762
Animals with one parent unknown	241	98
Number of founders	1,626	860
Effective number of founders	38.6	39.5
Effective number of ancestors	19	13
Number of ancestors explaining 100% genetic variability of population	994	417
Number of ancestors explaining 50% genetic variability of population	7	5

The genetic contribution of the FC founders was summed up by country of origin to quantify the importance of each country as a source of genes in the formation of the Spanish Arab Horse (Figure 1, see Plot B). FC founders with its country of origin identified with certainty contributed to the genetic variability of the population in a large extent (86.0%) whilst the contribution of the SN founders was of 9.9%.

Geographical areas contributing significantly to the genetic variability of the population were 'Middle East' (36.4%), Poland (18.7%), the United Kingdom (11.9%) and Egypt (9.1%). However, other countries such as USA or France that contributed to the Spanish Arab Horse with a significant number of founder individuals summed up contributions lower than 3% (2.3% and 2.1%, respectively), due to the very small contribution of these founder animals.

The number of ancestors (founders or not) that explained 100% of genetic variability of the whole population was 994, although only 7 individuals were necessary to explain 50% (see Table 2). For the reference population these figures were respectively 417 and 5. The effective number of ancestors was 19 and 13 respectively for the whole and the reference populations. Table 1 gives the sum of the contributions of the identified Spanish Arab ancestors according to their decade of birth. Major contributions were given by ancestors born during the decades 1895-1904 (12.5%), 1925-1934 (38.2%) and 1935-1944 (22.1%) whilst those born in the other periods have a sum of the contributions lower than 8%. This characterises the dramatic scenario suffered by the population during and after the Spanish Civil War period in which a few number of individuals were massively used for reproduction.

Table 3 shows the contribution of the 15 main ancestors (identified according to Boichard *et al.*, 1997) identified for the whole and the reference populations. Ancestors with the higher contribution to the Spanish Arab Horse were *Gandhy* (13.8% and 16.8%, respectively, for the whole and the reference populations) and *Congo* (12.7% and 16.3%, respectively, for the whole and the reference populations).

Table 3. Description of the 15 main ancestors (identified according to Boichard *et al.*, 1997) contributing the most to the Spanish Arab Horse population as a whole and in the last decade.

	Sex	Origin country	Year of Birth	Genetic contribution (%)
<i>Whole population</i>				
1. Gandhi	M	Spain	1931	13.8
2. Congo	M	Spain	1941	12.7
3. Nana-sahib	M	Spain	1934	7.2
4. Wan-dick	M	Poland	1901	6.7
5. Barquillo	M	Spain	1938	4.1
6. Yaima	F	Spain	1935	3.4
7. Sara 1930	F	Spain	1930	3.4
8. Veracruz	F	Spain	1934	2.8
9. Razada	M	England	1925	2.0
10. Korosko	M	Egypt	1907	1.9
11. Seanderich	M	Middle East	1902	1.8
12. Veranda	F	Spain	1933	1.7
13. Reyna	F	England	1925	1.6
14. Sabat-el-heir	M	Egypt	1906	1.5
15. Ursus	M	Poland	1908	1.4
<i>Reference Population</i>				
1. Gandhi	M	Spain	1931	16.8
2. Congo	M	Spain	1941	16.3
3. Nana-sahib	M	Spain	1934	7.8
4. Barquillo	M	Spain	1938	6.7
5. Yaima	F	Spain	1935	5.3
6. Sara 1930	F	Spain	1930	4.3
7. Veracruz	F	Spain	1934	3.9
8. Ocalina	F	Spain	1951	2.7
9. Abisinia 1935	F	Spain	1935	2.3
10. Wan-dick	M	Poland	1901	2.0
11. Veranda	F	Spain	1933	1.9
12. Nazeer	M	Egypt	1934	1.5
13. Saboya	F	Spain	1931	1.4
14. Ghazel	F	Middle East	1924	1.2
15. Kashmir	M	Spain	1934	1.1

Figure 3 shows the contribution of the historically important stallions (sire lines) to the population per decade of birth of their descendants.

The stallion *Ursus* was the most influential with an increasing importance over decades to finally reach values of 14.0% in the animals born during the last decade. The stallions *Seanderich* and *Razada* contributed moderately to the population (7.2% and 5.0%, respectively, during the last decade). Finally, *Bagdad* has shown to have a low genetic influence on the registered individuals (over 2% from 1945 to present).

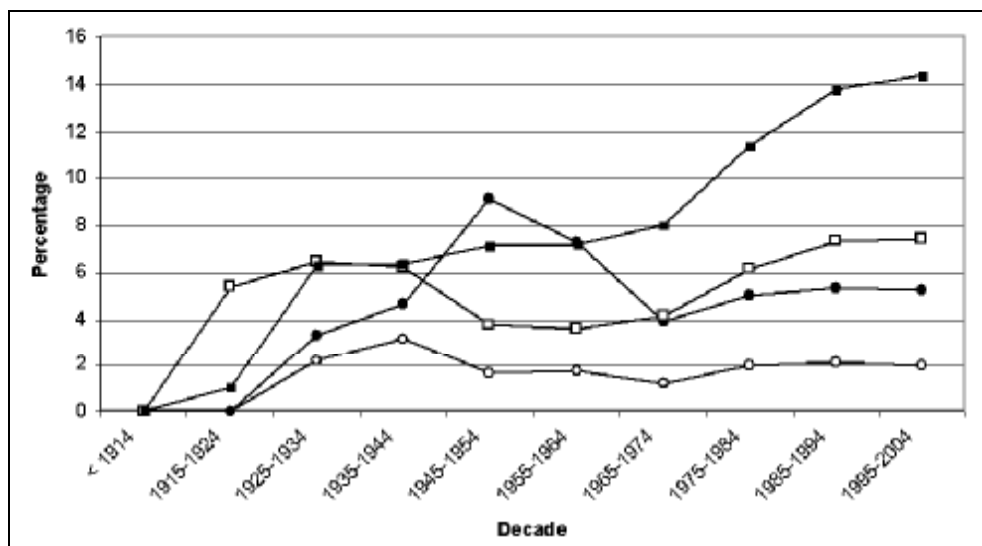


Figure 3. Influence of the sire lines *Ursus* (black squares), *Seanderich* (open squares), *Razada* (black circles) y *Bagdad* (open circles) in the Spanish Arab Horse by decade.

3.3. Inbreeding and Average Relatedness Coefficients

The average values of F and AR for the whole analysed pedigree were, respectively, 7.0% and 9.1%. The variation of average F and AR values by decades of birth of the animals is given in Table 1. The average F values were higher than 1% from the 10's being higher than 6% from 1975. This parameter reached average values of 9.8% in the reference population. During the last decade the number of inbred individuals is important. Table 4 gives the frequency of Spanish Arab Horse individuals per degree of inbreeding. Individuals with very high levels of inbreeding ($F \geq 12.5\%$) were 17.7% in the whole population and more than a quarter (26.7%) of the registered individuals born in the last decade. The variation of the mean AR values was similar to that of F with maximum values in the last decade of 11.6%. In order to distinguish between close and remote inbreeding the variation of the inbreeding by decades of birth computed using only data from the last 5 (F_5) generations was also given (Table 1). Patterns of variation of F and F_5 were consistent till the decade 1955-1964 when they reached values around 3%. After this decade, the average F_5 of the registered individuals

increased more slowly than F to reach maximum values of 6.1% in 1985-1994 whereas in the last decade decreased to 5.2%. However the maximum value of F was 9.8% for the same period.

Table 4. Absolute and relative (in percentage) frequency of individuals registered in Spanish Arab Horse by rate of inbreeding (F). Data are shown by the whole data set and for the last decade.

Rate of inbreeding	Whole population		Decade 1995-2004	
	N	%	N	%
Non inbred ($F = 0$)	5,007	26.5	550	8.9
$0 < F < 6.25\%$	3,709	19.6	956	15.4
$6.25\% \leq F < 12.5\%$	6,817	36.1	3,048	49.0
$F \geq 12.5\%$	3,347	17.7	1,660	26.7

3.4. Within population subdivision

Within population genetic differentiation was assessed using F -statistics computed from genealogies. This analysis was carried out at two different population levels considering: a) the sire line level; and b) the geographical area level (see material and methods section). At the sire line level the statistics F_{ST} reached values of $F_{ST} = 0.010$, $F_{IS} = 0.016$ and $F_{IT} = 0.025$. However, no differentiation was assessed by geographical area with values of the corresponding F -statistics of $F_{ST} = 0.001$, $F_{IS} = 0.019$ and $F_{IT} = 0.025$.

4. Discussion

4.1. Demographic analyses

The Arab horse is widely spread in the whole world founding particular populations in many countries. The Spanish Arab Horse was mainly formed using individuals from foreign countries and the studbook remains open today for FC Arab individuals.

The sum of the genetic contributions of the founder individuals to the population according to the country of origin makes possible to ascertain major sources of genes in the history of the Spanish Arab Horse. Here we show that the major source of genes of the Spanish Arab Horse is 'Middle East' coinciding with the historical major breeding area of the breed. Moreover, the importance of the Arab horse populations of Poland, the United Kingdom, and Egypt in the formation of the Spanish Arab has been also confirmed. Major genetic contributions by country of origin assessed here roughly coincide with the type of animals bred in the Spanish Arab studs (Fenaux, 1995): traditionally, they select individuals of 'Egyptian type' to improve type traits whilst individuals of 'Polish type' are preferred to improve movements.

The Spanish Arab Horse is a population with historical pedigrees that can be traced back to the mid 19th century (Maxwell, 1995). In this respect, the whole

Spanish Arab Horse pedigree has lower level of completeness than other comparable breeds due to the constant inclusion of FC individuals in the studbook. The Spanish Arab Horse, as a whole, had a completeness level higher than 74% when up to four generations are traced back. The Andalusian horse breed, with comparable pedigree depth, had completeness about 90% until this generation (Valera *et al.*, 2005). In Lipizzan horse a level of completeness of 90% was observed until the tenth generation (Zechner *et al.*, 2002). This relative lack of pedigree information has also been highlighted by the number of equivalent generations which is, for the whole Spanish Arab Horse population (5.7), lower than the values reported in Lipizzan (15.2; Zechner *et al.*, 2002) or Andalusian (8.3; Valera *et al.*, 2005). However, the quality of the available genealogical information has substantially increased during the last decades in which the interest of the breeders and the number of registrations in the studbook has increased significantly. In this respect, in our reference population the level of completeness is higher than 90% until sixth generation, similar to the value for the French Arab, over 90% in the fourth generation (Moureaux *et al.*, 1996).

4.2. Genetic variability

The assessment of the genetic variability of a population is a necessary step before the implementation of an improvement program. This will be determined by the size of the base population but also by the mating policy and population structure that will allow different degrees of genetic contributions from founder to the present population. The number of founders in the reference population of the Spanish Arab Horse (860) is lower to that previously reported for the French Arab (962 founders; Moureaux *et al.*, 1996). However, this kind of comparisons must be interpreted with caution since the ratio number of founders over total number of animals depends on the time from foundation of the Studbook and the quality of pedigree data recording. The total number of founders (1,626) of the Spanish Arab Horse population (that are 8.6% of the registered individuals) is similar to that recently reported for the Andalusian horse breed (1,465 founders) though in the later breed the founders account for roughly a 2% of the total population (Valera *et al.*, 2005). This is due to the main characteristic of the Spanish Arab studbook that remains open to other Arab horse population from its foundation to the present.

However, parameters characterising the genetic contributions from founders to the population do not reach higher values thus reflecting an unbalanced use of some individuals for reproduction which affects the whole genetic stock. The effective number of founders, which measures the overall founder representation in a managed population accounting for the loss of genetic variability from unequal founder contributions (Lacy, 1989, 1995) was 39.5 for the reference population in Spanish Arab Horse (4.6% of total founders), which is lower than that of 135 (14.0% of total founders) obtained in French Arab Horse (Moureaux *et al.*, 1996) and other French horse breeds such as Trotteur français (70; 2.8% of total founders) or Selle français (333; 3.7% of the founders) as reported by Moureaux *et al.* (1996). However, the analyses by Moureaux *et al.* (1996) differ from these presented here in levels of completeness, total number of founders and number of generations used for the study, and, in consequence, comparisons are not

straightforward. In any case, the effective number of founders for the whole population (38.6, 2.4% of total number of founders) in the Spanish Arab Horse is also lower than others recently reported in horse breeds with comparable pedigree depths such as Lipizzan (48.2; 10.55% of total founders; Zechner *et al.*, 2002) or Andalusian (39.6; 2.7%; Valera *et al.*, 2005) but higher than reported for Thoroughbred (28; 17.8%; Cunningham *et al.*, 2001). The effective number of ancestors (which is here 19 for the whole population, roughly half of the effective number of founders) that accounts for bottlenecks in the population confirms the abusive use of a small number of individuals for reproduction in the Spanish Arab Horse. Its value in Spanish Arab Horse is lower than that of 26.2 obtained for the whole pedigree in Lipizzan horses (Zechner *et al.*, 2002) and similar than that of 16.5 obtained for the whole pedigree in Andalusian horse (Valera *et al.*, 2005) that have more complete and deeper genealogies registered in the studbook.

Information given in Tables 2 and 3 is useful to characterise the breeding policy of the population. The individual contribution to the genetic variability of the population from the individuals identified as ancestors (Boichard *et al.*, 1997) are considered; the genetic contribution of the major two Spanish Arab ancestors to the whole population (13.8% and 12.7%) is quite similar to that reported in other horse breeds with unbalanced mating policies such as Andalusian horse (the major two ancestors contributing 15.8% and 12.6% respectively). In fact, only 7 Spanish Arab ancestors are needed to explain 50% of the genetic variability of the population whilst in the Andalusian horse breed and in the Lipizzan population 6 and 10 ancestors, respectively are needed (Valera *et al.*, 2005; Zechner *et al.*, 2002). The overall information reported here characterises the Spanish Arab horse as a population that, regardless the high number of founders registered in the studbook, has followed a mating policy allowing for the very unbalanced use of some individuals. This can have an historical reason that can arise from the genetic bottleneck suffered by the population due to the Spanish Civil War (Table 1). Genetic contributions of the ancestors born during the period of the War and the previous decade summed up approximately a 50%. Most major ancestors described in Table 3 were born during these decades and, in fact, we can consider that the Spanish Arab Horse was genetically 're-founded' during the late 30's and the 40's.

4.3. Inbreeding

The average F value computed for the whole pedigree of Spanish Arab Horse population is lower than that computed in other Spanish populations as Andalusian horse (8.5%; Valera *et al.*, 2005). This can be explained by the special scenario described in our Arab horse in which the studbook is still open for the registration of Arab individuals. The pedigree of the new registered individuals is not always completely included in the Spanish Arab Studbook, producing subsequently animals considered as founders. However, we had expected lower levels of inbreeding than those assessed in this scenario in which the rate between the number of founder individuals and the total population size exceed that of breeds with comparable pedigree depth. The average F computed using the reference population (9.8%) is higher than those of 3.1% and 5.3% reported, respectively, by

Moureaux *et al.* (1996) and Głażewska and Jezierski, (2004) in French and Polish Arab horses, for the individuals born during the more recent period in their databases. This fact and the proportion of individuals with levels of inbreeding higher than 12.5% (see Table 4) point out that breeders have planned most matings using related individuals. Recently, Głażewska and Gralak, (2006) have shown in Polish Arab horse that breeders have carried out a balancing selection favouring the more heterozygous mares (at the protein polymorphism level) for reproduction. This can cause discrepancies between real gene pool in a population and that estimated on the basis of pedigrees (Głażewska and Gralak, 2006; Royo *et al.*, 2007).

Moreover, after 1975 average F is higher than half mean AR thus pointing out that matings are mainly carried out between relatives. This is also reflected by the parameter F_{IS} . The F_{IS} statistic computed from genealogies is equivalent to the parameter α (Caballero and Toro, 2000; Toro *et al.*, 2000), which means the departure from random mating as a deviation from Hardy-Weinberg proportions. Positive genealogical F_{IS} values mean that the average F within a subpopulation exceeds between-individuals coancestry (Gutiérrez *et al.*, 2005) and occurs if mating between relatives are not avoided (Toro *et al.*, 2000). F_{IS} values computed at the sire line level and at geographical area level reached positive values near 2% (0.016 and 0.019, respectively). However, this breeding policy might have been changed these past few years. The trends of the inbreeding computed using only the last 5 generations (F_5) showed that during the last decade the average F_5 was roughly 1% lower than that computed in the ante last decade. F depends to a large extent on the quality of pedigree information especially in the initial generations registered. F_5 still increases in our Arab horse population up to 1994 accounting for the mating policy described above. However in the last decade a major effort seems to have been made to correct the possible effect of such high levels of inbreeding in the population.

The wide use of particular individuals in the population is also reflected in the lack of within-population genetic differentiation assessed using genealogical F_{ST} 's. At the sire line level a little differentiation over 1% was obtained due to the predominance of some sire genealogies in particular studs. However, stallions assigned to a given paternal line share genes (inherited by their remaining ancestors) with the other sire lines. So, traditional assignment of a stallion to a sire line in the Spanish Arab Horse (Maxwell, 1995) does not have clear genetic support. Moreover, there is no geographical structure in our Arab horse. The expansion of the breed from Andalusia to other parts of Spain has not been made by using particular sources of genes; neither from FC nor SN individuals. The breeding significance of the Military Stud of Jerez de la Frontera has probably caused this lack of geographic differentiation. In fact, the Military stud continues to be the most significant source of genes in the Spanish Arab Horse. This governmental stud sets up public mounting services throughout the Spanish territory using its stallions those that are used in most medium-small studs.

5. Conclusions

Here we carried out the first analysis of the pedigree information registered in the Spanish Arab Horse studbook to assess the genetic variability and the population structure. The information obtained will be useful to implement a selection program in the Spanish Arab Horse. The high number of founders registered in the studbook would suggest the existence of a large genetic basis for a stallions selection programme. However, historical constraints, such as the consequences of the Spanish Civil War, and the breeding policy carried out by the breeders has reduced the available genetic variability as reflected in small values computed for parameters such as the effective number of founders and the effective number of ancestors. In the light of our results a selection programme in the Spanish Arab Horse would need an appropriate management of the genetic stock. The implementation of mating policies promoting the unbalanced use of those individuals with lower *AR* values for reproduction, as proposed previously for other Spanish horse breeds (Valera *et al.*, 2005) and the exchanges between Arab populations from different countries can be useful to maintain the genetic variability of the population and also to enlarge genetic basis for selection.

Acknowledgements

The authors thank the support given by the 'Subdirección General de Medios de Producción Ganaderos' of the Spanish Ministry of Agriculture. This work has been made within a collaboration agreement between the Spanish Arab horse breeders association (AECCA) and the University of Sevilla. We also thank the 'Fondo de Explotación de los Servicios de Cría Caballar y Remonta' of the Spanish Ministry of Defence for its collaboration. This work was partially funded by a grant from the Junta de Andalucía given to I. Cervantes (BOJA, nº. 120, 21/06/2004).

References

- Boichard, D., Maignel, L., Verrier, E., 1997. The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Sel. Evol.* 29, 5-23.
- Bowling, A.T., Ruvinsky, A., 2000a. The genetic resources and their conservation. In. "The Genetic of the horse". CAB International. UK.
- Bowling, A.T., Ruvinsky, A., 2000b. Genetic Aspects of Domestication. In. "The Genetic of the horse". CAB International. UK.
- Caballero, A., Toro, M.A., 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genet. Res. Camb.* 75, 331-343.
- Caballero, A.; Toro, M.A., 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv. Genet.* 3, 289-299.
- Cunningham, E. P., Dooley, J.J., Splan, R.K., Bradley, D.G., 2001. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Anim. Genet.* 32, 360-364.
- Falconer, D.S., Mackay, F.C., 1996. Introduction to Quantitative Genetics. Longman Group Ltd. England.

- Fenau, K., 1995. El Pura Sangre Árabe. Ediciones el Caballo, Barcelona. España.
- Głazewska, I., Jezierski, T., 2004. Pedigree analysis of Polish Arabian Horses based on founder contributions. *Livest. Prod. Sci.* 90, 293-298.
- Głazewska, I., Gralak, B., 2006. Balancing selection in Polish Arabian Horses. *Livest. Sci.* 105, 272-276.
- Goyache, F., Gutiérrez, J.P., Fernández, I., Gómez, J., Álvarez, I., Diez, J., Royo, L.J., 2003. Using pedigree information to monitor genetic variability of endangered population: the Xalda sheep breed of Asturias as an example. *J. Anim. Breed. Genet.* 120, 95-103.
- Gutiérrez, J.P., Altarriba, J., Díaz, C., Quintanilla, R., Cañón, J., Piedrafita, J., 2003. Genetic analysis of eight Spanish beef cattle breeds. *Genet. Sel. Evol.* 35, 43-64.
- Gutiérrez, J.P., Goyache, F., 2005. A note on ENDOG: a computer program for analysing pedigree information. *J. Anim. Breed. Genet.* 122, 172-176.
- Gutiérrez, J.P., Marmí, J., Goyache, F., Jordana, J., 2005. Pedigree information reveals moderate to high levels of inbreeding and a population genetic structure in the Catalanian donkey breed. *J. Anim. Breed. Genet.* 122, 378-386.
- James, J., 1972. Computation of genetic contributions from pedigrees. *Theor. Appl. Genet.* 42, 272-273.
- Lacy, R.C., 1989. Analysis of Founder Representation in Pedigrees: Founder Equivalent and Founder Genome Equivalents. *Zoo. Biol.* 8, 111-123.
- Lacy, R.C., 1995. Clarification of the genetic terms and their use in the management of captive populations. *Zoo. Biol.* 14, 565-578.
- MacCluer, J., Boyce, B., Buke, L., Weitzkamp, D., Pfenning, A., parsons, C., 1983. Inbreeding and pedigree structure in Standardbred horses. *J. Heredity* 74, 394-399.
- Malécot, G. 1948. Les Mathématiques de l'Hérédité. Masson et Cie, Paris.
- Maxwell, J., 1995. Spanish Arabian Horse Families. Alexander Heriot & Co. Ltd, England.
- Moureaux, S., Verrier, E., Ricard, A., Meriaux, J.C., 1996. Genetic variability within French race and riding horse breeds from genealogical data and blood marker polymorphisms. *Genet. Sel. Evol.* 28, 83-102.
- MAPyA., 2003. Estudio y caracterización genética del sector equino en España. Ministerio de Agricultura Pesca y Alimentación.
(<http://www.mapa.es/app/Equino/Informacion/Infsector.aspx?lng=es>). España.
- Royo, L.J., Álvarez, I., Gutiérrez, J.P., Fernández, I., Goyache, F., 2007. Genetic variability in the endangered Asturcón pony assessed using genealogical and molecular information. *Livest. Sci.* 107, 162-169.
- Steen, A.K., 2006. A glorious task. A tale of the breed. Escandón Impresores. Seville. Spain.
- Toro, M.A., Rodrigáñez, J., Silió, M., Rodríguez, C., 2000. Genealogical analysis of a close herd Black Hairless Iberian pigs. *Conserv. Biol.* 14, 1843-1851.
- Valera, M., Molina, A., Gutiérrez, J.P., Gómez, J., Goyache, F., 2005. Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain. *Livest. Prod. Sci.* 95, 57-66.

Wright, S., 1978. Evolution and the genetics of populations: Vol. 4. Variability within and among natural populations. University of Chicago Press: Chicago. USA

Zechner, P., Sölkner, J., Bodo, I., Druml, T., Baumung, R., Achmann, R., Marti, E., Habe, F., Brem, G., 2002. Analysis Diversity and population structure in the Lipizzan horse bred based on pedigree information. Livest. Prod. Sci. 77, 137-146.

4.2. ARTICULO 2: GENEALOGICAL ANALYSES IN OPEN POPULATIONS: THE CASE OF THREE ARAB-DERIVED SPANISH HORSE BREEDS

I. Cervantes^{1,2*}, J.P. Gutiérrez¹, A. Molina², F. Goyache³ and M. Valera⁴

¹Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro s/n, E-28040-Madrid, Spain.

²Departamento de Genética, Universidad de Córdoba, Ctra. Madrid-Cádiz, km 396^a, E-14071 Córdoba, Spain.

³SERIDA-Somío, C/ Camino de los Claveles 604, E-33203 Gijón (Asturias), Spain.

⁴Departamento de Ciencias Agro-Forestales, EUITA, Universidad de Sevilla, Ctra. Utrera km 1, E-41013 Sevilla, Spain.

*Correspondence: I. Cervantes, Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro s/n, E-28040-Madrid, Spain. Phone and Fax: +34913943767. E-mail: icervantes@vet.ucm.es

Summary

This research was to assess the genetic composition of three Arab-derived Spanish horse breeds as an example to highlight the major shortcomings related to genealogical analyses in open populations and to propose approaches useful to deal with this task. The studbooks of three Spanish Arab (SA) derived horse breeds, Spanish Anglo-Arab (dAA), Hispano-Arab (dHA) and Spanish Sport Horse (dSSH) and those of their parental breeds SA, Spanish Purebred (SPB) and Thoroughbred (TB), totaling 211,754 individuals, were available. The genealogies of the dAA, dHA and dSSH were analysed using the corresponding studbook (breed exclusive dataset) but also including the genealogies of the founders from parental breeds (completed dataset). The effective number of founders and ancestors computed from the completed datasets were reduced of those obtained from the breed exclusive pedigrees in the dAA, dHA and dSSH breeds. Coancestry analyses revealed that the present SA-derived populations share more genes with the Arab than with the other parental breeds. Effective population size was computed by accounting for migration rates to obtain an equivalent closed-population effective size ($_{eq}N_e$) of 39.2 for the dAA, 56.3 for dHA and 114.1 for dSSH. The essayed methodologies were useful for characterising populations involving migration. The consequences of the management of the analysed breeds are discussed.

Key words: effective population size; migration; open populations; pedigree analysis; Spanish Arab horse.

Introduction

The Spanish Arab (SA) horse breed studbook was founded in 1847 by the Spanish Ministry of War (Maxwell, 1995). Historically, crossbreeding to obtain individuals with the desired performance characteristics was the major purpose of the Arab horses in Spain (Maxwell, 1995). This is the basis of the development of a number of Spanish Arab-derived horse breeds. Two major Spanish horse breeds, Anglo-Arab (dAA) and Hispano-Arab (dHA) constituting 5.3% of the national equine census (MAPyA, 2003), were formed by continuous cross breeding between SA horses and Thoroughbred (TB) and Spanish Purebred (SPB; Andalusian) horses respectively. Their studbooks were founded respectively in 1884 and 1986, remaining open to individuals with an Arab parent. The Anglo-Arab is one of the most ancient Spanish Studbooks. Despite this there is a consensus on the existence of the Hispano-Arab horse population early in the 20th century (Steen, 2006), genealogical recording begun recently with the foundation of its Studbook. This breed in Spain is officially considered as an endangered breed (Spanish regulation RD 1662/1997). Additionally, the development of the Spanish equine industry (Valera *et al.*, 2005; Azor *et al.*, 2006; Cervantes *et al.*, 2008a) stimulated by the increasing use of horses for sports and recreation, has led to the creation of a new breed, the Spanish Sport horse (dSSH; Studbook founded in 2002), with some genetic influence from the SA breed. Despite the importance of this scenario, no analysis on the genetic influence of the SA horse on the Arab-derived breeds has been carried out. Furthermore, the problems regarding to inbreeding, population size or crossbreeding in Spanish Arab-derived breeds were unknown until this study.

The genealogical analyses of open populations, allowing the use of individuals from other breeds for reproduction, present particular characteristics that make it

difficult to ascertain the real rates of genetic variability in the analysed stock (Hamann and Distl, 2008). The analysis of populations receiving genetic contributions from other populations over time needs to account for migration rates to assess their genetic variability in terms of inbreeding and effective size (Falconer and MacKay, 1996). Moreover, the effect of the mating policy on the genetic variability of an open population cannot be straightforwardly assessed via the computation of the effective population size because this parameter is influenced by rates of migration from external populations.

In this respect, in order to ascertain the genetic influence of the source populations in terms of founder contributions one needs to trace genealogies back using pedigrees from pure studbooks of origin to identify founders with certainty. However, the impact of ignoring those genealogies from pure breeds is usually unknown (Hamann and Distl, 2008).

The aim of this research was to assess the genetic composition of three SA-derived Spanish horse breeds and the relationships and the differentiation between them using genealogical information, addressing the consequences of ignoring the parental pedigrees, as well as developing and essaying methodologies to manage populations submitted for migration. To deal with these objectives the complete Spanish horse studbooks for the SA, SPB, TB, dHA, dAA and dSSH horse breeds were used. The consequences of the obtained results for the management of the analysed breeds are discussed.

Material and methods

Pedigree data

Information registered in the studbooks (from their foundation to December 2004) of the Spanish Anglo-Arab, Hispano-Arab and Spanish Sport horse breeds, as well as the full genealogies of the Spanish Arab, Spanish Purebred and Spanish Thoroughbred horse breeds, were obtained from the official web site of the Spanish horse studbook (Ministry of Defense: www.librogenealogico.com), and used in this analysis. The total number of animals was: 8,289 for dAA, 3,394 for dHA, 7,099 for dSSH, 18,880 for SA, 140,629 for SPB and 33,463 for TB. The records of animals born per period of five years from the dAA, dHA and dSSH studbooks are summarized in Table 1.

Table 1. Number of available records included in the Spanish Anglo-Arab (dAA), Hispano-Arab (dHA) and Spanish Sport horse (dSSH) studbooks per 5 year periods.

	dAA	dHA	dSSH		dAA	dHA	dSSH
<1925	445			1965-1969	436	7	2
1925-1929	79			1970-1974	515	28	13
1930-1934	83			1975-1979	680	91	24
1935-1939	43			1980-1984	588	214	149
1940-1944	111			1985-1989	503	393	226
1945-1949	245			1990-1994	820	874	624
1950-1954	383	3		1995-1999	1286	921	1618
1955-1959	261			2000-2004	1382	862	4437
1960-1964	429	1	6	total	8289	3394	7099

Note that a given SA individual can act as parent in each of the dAA, dHA and dSSH studbooks. Furthermore, dSSH is a composite breed that includes in its studbook any individual participating in sports competitions and that is not registered in any other studbook and, therefore, a given SA, SPB, TB, dAA or dHA individual can act as parent in the dSSH studbook. The available information was edited to avoid redundancies and to make the identification of the same individual uniform regardless of the breed exclusive database in which it is recorded and to identify with certainty the breed of each individual.

Throughout the manuscript the dAA, dHA and dSSH breeds will be referred to as 'derived' populations whilst the SA, SPB and TB breeds will be considered 'parental' populations.

Table 2 shows the number of individuals acting as parents in the breed exclusive datasets (those directly obtained from the SA-derived studbooks) of those derived by parental horse breeds. A total of 21.4% of individuals acted as parents were SA and 26.3% were TB in the dAA studbook; a total of 28.3% for SA and 42.2% for SPB individuals acted as parents in the dHA studbook; and finally, a total of 4.4% SA, 19.1% SPB, 10.3% TB, 5.3% dAA, 1.6% dHA and 36.6% of individuals from other breeds acted as parents in the dSSH studbook.

Within pedigrees, a group of animals approaching the last generation were defined as 'reference populations' within each population. For this purpose, the average generation length (GL), defined as the average age of parents at the birth of their offspring kept for reproduction, was computed for each derived population using the program ENDOG (Gutiérrez and Goyache, 2005). According to average GLs (12.0 ± 0.13 years for dAA, 10.1 ± 0.16 years for dHA and 11.4 ± 0.21 years for dSSH) the reference populations finally included the animals born in the last 12 years for the dAA breed, 10 years for dHA, and 11 years for dSSH.

Table 2 also gives the number of stallions and mares acting as parents of the reference populations. In dAA the highest proportion of mares (46.8%) were dAA whilst for stallions these were SA (44.7%); in dHA, the highest proportion of mares (50.4%) were dHA and for stallions these were SPB (66.3%); with respect to dSSH, the highest proportion of mares were dSSH (36.7%) and for stallions these were SPB (39.9%).

The pedigree information summarized in Table 2 was used to include all the available genealogies of the parents identified from the SA, SPB and TB studbooks in the dAA, dHA and dSSH datasets. The datasets were completed with parental genealogies given that, in the breed exclusive datasets, the genealogies were truncated where an ancestor from a parental breed appeared. After including information from the parental studbooks the available records totaled 16,304, 10,605, and 21,826, respectively, for dAA, dHA and dSSH. Throughout the manuscript the information contained in dAA, dHA and dSSH studbooks without recording 'parental genealogies', will be referred to as 'breed exclusive datasets' and after including the information from the parental studbooks they will be referred to as 'completed datasets'. Note that, regardless of whether the analysed dataset was breed exclusive or completed, those individuals with both parents unknown were considered as founders following Moureaux *et al.* (1996), Valera *et al.* (2005) and Cervantes *et al.* (2008a). As a consequence, most individuals that were considered founders in the breed exclusive datasets were not considered as such in the completed datasets. Note also that the number of animals acting as parents, in the reference populations are defined the same before and after completing the genealogies. Additionally, according to Lacy (1989) and Cervantes *et al.* (2008a) when only one parent is known, the unknown parent was considered a founder.

The contributions of founders within breed of origin were summed up to calculate the influence of the different pure breeds. When a founder appeared assigned to its own studbook instead a parental breed, it was assumed that it had a parental composition equal to the average contribution of each parental population to the corresponding derived breed.

Table 2. Number of individuals acting as parents in the breed exclusive datasets of the Spanish Anglo-Arab (dAA), Hispano-Arab (dHA) and Spanish-Sport horse (dSSH) studbooks, by parental horse breed. Whole pedigrees included all animals in the dataset and reference populations included animals born in the last 12 years for the dAA breed, 10 years for dHA, and 11 years for dSSH. In brackets, % with respect to the total parents, total stallions and total mares, respectively.

Parental Breed	Whole pedigrees			Reference populations					
	dAA	dHA	dSSH	dAA		dHA		dSSH	
				Stallion	Mare	Stallion	Mare	Stallion	Mare
Spanish Arab	1,016	644	255	346	225	182	269	171	47
(SA)	(21.4%)	(28.3%)	(4.4%)	(44.7%)	(15.1%)	(25.5%)	(30.3%)	(9.7%)	(1.5%)
Spanish Purebred		959	1,102			474	172	703	278
(SPB)		(42.2%)	(19.1%)			(66.3%)	(19.3%)	(39.9%)	(8.9%)
Thoroughbred	1,253		595	166	567			156	315
(TB)	(26.3%)		(10.3%)	(21.5%)	(38.1%)			(8.8%)	(10.1%)
Anglo-Arab	2,486		307	261	698			103	149
(dAA)	(52.3%)		(5.3%)	(33.8%)	(46.8%)			(5.8%)	(4.8%)
Hispano-Arab		669	91			59	448	26	61
(dHA)		(29.5%)	(1.6%)			(8.2%)	(50.4%)	(1.5%)	(2.0%)
Spanish-Sport			1,301					203	1144
(dSSH)			(22.6%)					(11.5%)	(36.7%)
Others ^a			2,110					400	1122
			(36.6%)					(22.7%)	(36.0%)

^aincludes all parents belonging to non-listed breeds

Statistical analyses

A number of population and genetic parameters were computed on the reference populations using the program ENDOG (current version v4.5; Gutiérrez and Goyache, 2005).

The pedigree completeness level was computed for both the breed exclusive and the completed datasets as the proportion of ancestors known per parental generation (MacCluer *et al.*, 1983). Also, the number of equivalent to discrete generations (t) for each individual in a pedigree was computed as the sum of $(\frac{1}{2})^n$ where n is the number of generations separating the individual to each known ancestor (Boichard *et al.*, 1997).

The inbreeding coefficient (F), defined as the probability that two alleles at a randomly chosen locus are identical by descent (Malécot, 1948), and the average relatedness coefficient (AR), defined as the probability that an allele randomly chosen from the whole population belongs to a given animal (Goyache *et al.*, 2003; Gutiérrez *et al.*, 2003), were computed for each individual included in the breed exclusive and the completed datasets. Note that the AR coefficient of a founder means its genetic contribution to the population. For each reference population, these coefficients can be summed up for the founders belonging to a parental breed to ascertain the relative contributions from parental breeds.

The probability of gene origin was characterised by computing the following parameters: a) effective number of founders (f_e), which is the reciprocal of the probability that two alleles drawn at random in the studied population originate from the same founder (James, 1972) and this is computed from the genetic contribution of founders to the descendant gene pool of the population (Lacy, 1989); b) effective number of ancestors (f_a), defined as the minimum number of ancestors, not necessarily founders, explaining the complete genetic diversity of a population (Boichard *et al.*, 1997). Parameter f_a does not fully account for gene loss by drift from the ancestors to a reference population but complements the information offered by f_e accounting for the losses of genetic variability produced by the unbalanced use of reproductive individuals producing bottlenecks (Gutiérrez *et al.*, 2005); c) the founder genome equivalents (Ballou and Lacy, 1995), defined as the theoretically expected number of founders that would be required to provide the genetic diversity in the actual population if the founders were equally represented and had lost no alleles, was obtained by the inverse of twice the average coancestry of the individuals within the population (Caballero and Toro, 2000).

Effective population size (N_e) was computed following Gutiérrez *et al.* (2003) approaching the increase in inbreeding (ΔF) as the regression coefficient (b) of the individual inbreeding coefficient over the equivalent discrete generations (t), and considering the corresponding regression coefficient as the increase in inbreeding

between two generations ($b = F_t - F_{t-1} \approx \frac{F_t - F_{t-1}}{1 - F_{t-1}}$), and consequently $N_e = 1/(2b)$.

The migration rate (m) was computed for the breed exclusive dataset as the ratio between the number of animals from the parental populations acting as parents over the whole number of parents (from the parental population or not) identified in the reference population.

Assuming that inbreeding remains stable as a consequence of the balance between migration and drift in open populations, the equivalent to closed-population effective size ($_{eq}N_e$) was approached in the breed exclusive datasets

from $F \cong \frac{1}{4m_{eq}N_e + 1}$ (Falconer and Mackay, 1996), where F is the average

inbreeding in the reference population in which mean inbreeding remains constant and m is the migration rate. Using this expression we assume that a particular derived breed behaves as a population under the island migration model (Wright, 1931). The parameter $_{eq}N_e$ can only therefore be computed when no significant regression of F over the year of birth for the animals existed in the reference population and can be interpreted as the N_e corresponding to a population with the same mating policy as the analysed population but without migration. Note that the

formula $F \cong \frac{1}{4m_{eq}N_e + 1}$ is just a case (assuming that mutation is negligible) of

the more general formula $F \cong \frac{1}{4_{eq}N_e(u + v + m) + 1}$ where u and v are the

mutation rates in two directions between two alleles at a locus. The expression was developed to study each subpopulation under the island model, assuming no relationships among migrants, and therefore applies for the last purpose to our breed exclusive datasets in which migrants appear as founders. Note again that the $_{eq}N_e$ would give information on the expected N_e of a population with the same breeding policy as that of the studied breed but without migration, which would allow comparisons of mating policies with other populations via N_e .

Finally, the genetic relationships and differentiation between all reference populations composed by the last generation of individuals belonging to the parental and the derived populations (SA, SPB, TB, dAA, dHA and dSSH) were assessed from a database that included all the records of the six available studbooks. Here the relationship between animals previously considered as migrants is now taken into account. The coefficient of coancestry of two individuals is defined as the probability that two gametes taken at random, one from each, carry alleles that are identical by descent (Falconer and Mackay, 1996). The within-breed coancestry (f_{ii}) and the between-breeds coancestry matrix (f_{ij}) were computed averaging all pairwise coancestry coefficients of the individuals belonging, respectively, to a given breed i or to two different breeds i and j . Following Caballero and Toro (2000, 2002) the between-breeds Nei's minimum distance (D_m) matrix was also computed as $D_m = [(f_{ii} + f_{jj})/2] - f_{ij}$ where f_{ii} and f_{jj} are the average coancestry within two breeds i and j and f_{ij} the coancestry between two breeds i and j .

Results

Pedigree completeness

The quality of the pedigree information of the individuals included in the reference population is illustrated in Figure 1. The inclusion of the available pedigrees obtained from the parental population pedigrees significantly increased the completeness of the genealogies of the dAA, dHA and dSSH breeds. The breed exclusive datasets had a completeness of 51.2% for the dAA breed, 30.3% for dHA, and 24.7% for dSSH in the second parental generation whilst these figures in the completed pedigrees were, respectively, 98.8%, 91.4% and 59.6%. The highest effect on pedigree knowledge was observed for the dHA breed because of the influence of the SPB genealogies: the completed dataset of the dHA breed has levels of completeness higher than 60% up to the eighth parental generation (Valera *et al.*, 2005). The average number of equivalent to complete discrete generation gave similar information. The breed exclusive datasets of the dAA, dHA and dSSH breeds had average t of 2.54 for dAA, 1.35 for dHA, and 1.23 for dSSH whilst the figures corresponding to the completed datasets were, respectively, 5.60, 7.71 and 3.35. Thus, the completed pedigrees of the dAA, dHA and dSSH breeds were 2.2, 5.7 and 2.7 times depth, respectively.

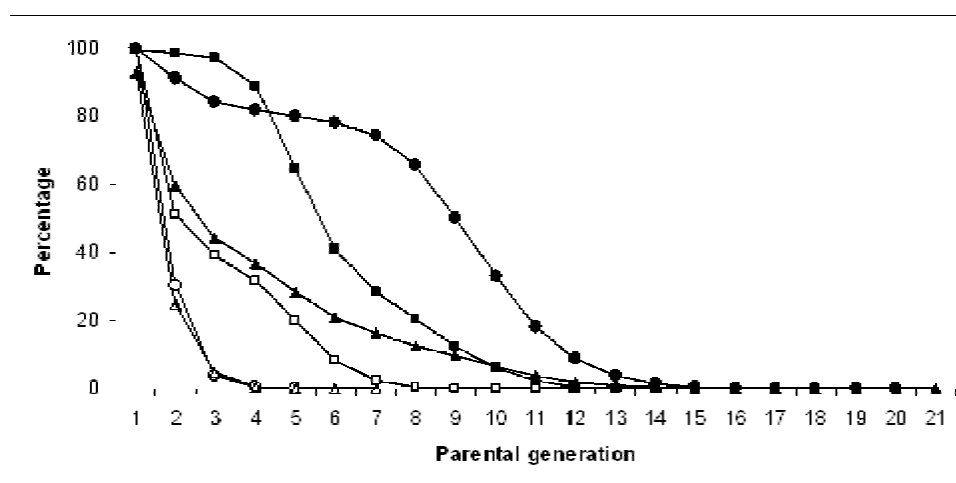


Figure 1. Percent of ancestors known per parental generation in Spanish Anglo-Arab (in squares), Hispano-Arab (in circles) and Spanish Sport horse (in triangles), with parental generation 1 that corresponds to parents, 2 that corresponds to grandparents, etc. Results from the breed exclusive datasets are in open symbols whilst those from the completed datasets after including information from the parental studbooks are in black symbols.

Probability of gene origin

Parameters characterising the genetic variability of the identified reference populations in terms of gene origin are given in Table 3. A total of 1,942, 1,417, 4,966 founders were identified in the breed exclusive datasets of the dAA, dHA and

dSSH breeds, respectively. However, the complete datasets allowed us to identify a different scenario with 2,414 founders for the dAA breed, 860 founders for dHA, and 5,655 founders for dSSH. The substantial decrease of total number of founders (39%) found in the completed pedigree of dHA, with respect to those identified in the breed exclusive pedigree, would characterise the dHA breed as a product of a limited number of families from the parental breeds thus limiting the genetic variability of the population.

The effective number of founders computed from the completed datasets were 48% for dAA, 27% for dHA, and 79% for dSSH of those obtained using the breed exclusive datasets. In turn, the effective number of ancestors computed from the completed datasets were 41% for dAA, 13% for dHA, and 59% for dSSH from those obtained using the breed exclusive datasets. And regarding the founder genome equivalents these values become 31% for dAA, 7% for dHA, and 39% for dSSH.

Table 3. Parameters characterising the probability of gene origin of the animals included in the reference populations, animals born in the last 12 years for the Spanish Anglo-Arab, 10 years for Hispano-Arab and 11 years for the Spanish-Sport horse breeds using the breed exclusive and the completed pedigree datasets.

	Spanish Anglo Arab		Hispano Arab		Spanish Sport Horse	
	breed exclusive ^a	completed ^b	breed exclusive ^a	completed ^b	breed exclusive ^a	completed ^b
Total number of animals in the reference population	3071	3071	1783	1783	6198	6198
Animals with both parents unknown	1883	2228	1377	670	4467	3889
Total number of founders	1942	2414	1417	860	4966	5655
Effective number of founders	433	210	327	87	762	602
Effective number of ancestors	241	98	311	41	736	436
Founder genomes equivalents	163	50	263	18	655	255
Ancestors explaining 100% of the genetic variability	1569	1202	1215	514	4272	4032
Ancestors explaining 50% of the genetic variability	130	44	203	16	638	284

^ausing breed exclusive dataset, ^busing completed dataset after including information from the parental studbooks

Table 4 details the contributions to the derived breeds of the founders belonging to the paternal breeds as identified in the completed datasets. The contributions of the Spanish Arab breed to its derived dAA and dHA breeds are lower than 50% and tended to be higher on the mares path. The lowest contribution of the SA founders was found in the dSSH breed (below 10%). In the dSSH breed the contributions of the founders belonging to the SPB and TB breeds, reached values

of roughly 20% each. However, most contributions (higher than 50%) were not from the populations assumed here to be parental: SA, SPB and TB.

Table 4. Contribution of the founders from the paternal populations to the derived Spanish Anglo-Arab (dAA), Hispano-Arab (dHA) and Spanish-Sport horse (dSSH) populations for the whole pedigree (WP, including all animals of the dataset) and for the reference populations (RP, including animals born in the last 12 years for the dAA breed, 10 years for dHA, and 11 years for dSSH) available from the breed exclusive datasets.

		Spanish Arab			Thoroughbred			Spanish Purebred			Others		
		Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total
dAA	WP	17.8	22.2	40.0	37.1	22.9	60.0						
	RP	24.7	16.7	41.4	30.6	28.0	58.6						
dHA	WP	20.4	26.2	46.6				33.5	19.9	53.4			
	RP	22.8	25.6	48.4				34.7	16.9	51.6			
dSSH	WP	6.1	2.9	9.0	10.0	9.0	19.0	14.9	6.0	20.9	19.3	31.8	51.1
	RP	6.1	2.7	8.8	10.9	8.1	19.0	13.2	4.8	18.0	25.2	29.0	54.2

Figure 2 shows the contributions of the founders of the parental breeds to the reference populations defined in the dAA, dHA and dSSH pedigrees by date of birth of the founders in order to illustrate the genetic history of the derived breeds. The SA influence on both the dAA and dHA breeds was always lower than that from TBs and SPBs, respectively. This was basically due to founders born between 1975 and 1994 (Plot 2A and 2B). The SA contribution to the dSSH breed (Plot 2C) was low and remained constant over the period analysed (over 2%). However, major founder contributions to the dSSH breeds came from breeds that were not included among the three parental populations analysed here.

Figure 3 shows the contributions of ancestors to the reference populations defined in the dAA (Plot 3A), dHA (Plot 3B) and dSSH (Plot 3C) pedigrees by date of birth of the ancestors. In dAA major genetic contributions came from ancestors belonging to their own derived breed (42.9%), with the contributions of the TB ancestors being higher (32.6%) than that from the SA ancestors (24.5%). In the dHA breed the ancestors belonging to the SA and SPB breeds gave similar genetic contributions (30.2% and 35.7% respectively). The contributions of parental ancestors to both the dAA and the dHA population were basically given by individuals born during the period 1985-1994. In dSSH, the genetic contributions of the ancestors belonging to the SA were stable and very low (3.3%). In this breed major contributions came from non-listed parental breeds and, increasingly, from dSSH ancestors (30.4%).

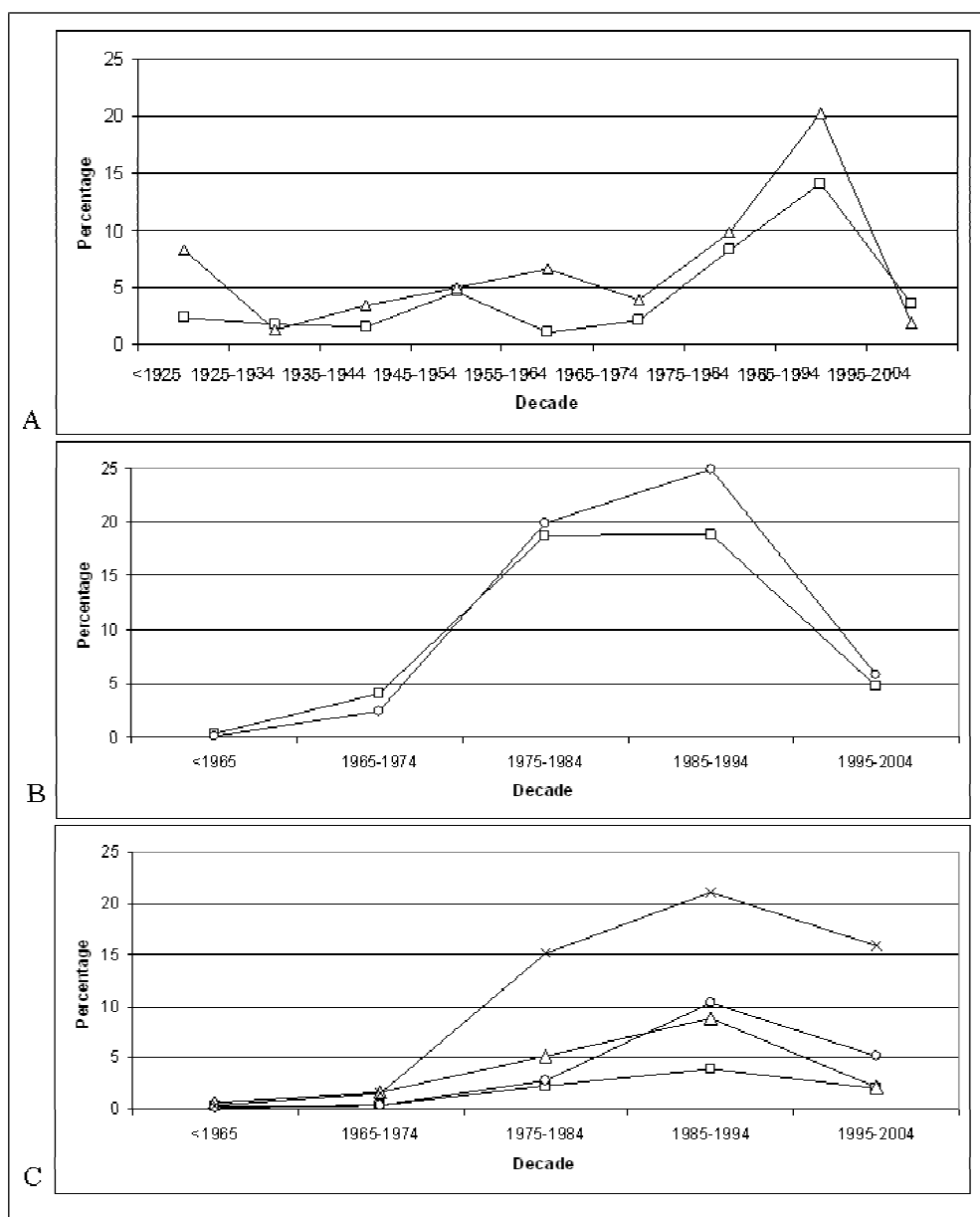


Figure 2. Contribution of the founders from the parental breeds to the reference populations (animals born in the last 12 years for the Spanish Anglo-Arab, 10 years for the Hispano-Arab, and 11 years for the Spanish Sport horse) of the Spanish Anglo-Arab breed (Plot A), Hispano-Arab breed (Plot B) and Spanish-Sport horse breed (Plot C), by decade of birth of the founders. Contributions of the Spanish Arab (□), Thoroughbred (Δ), Spanish Purebred (○) and from non-listed breeds (x) founders are represented.

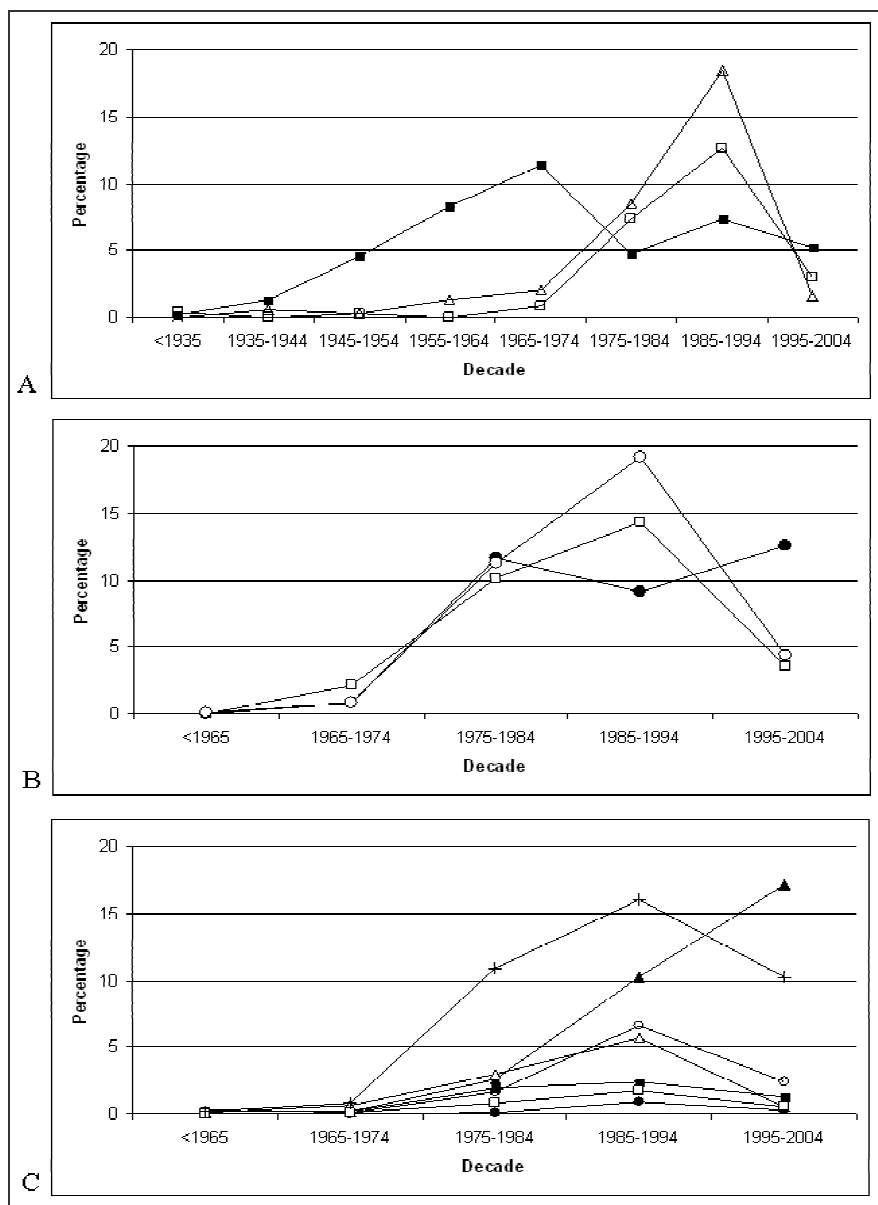


Figure 3. Contribution of the of Boichard *et al.*'s (1997) ancestors to the reference populations (animals born in the last 12 years for the Spanish Anglo-Arab, 10 years for the Hispano-Arab, and 11 years for the Spanish Sport horse) of the Spanish Anglo-Arab breed (Plot A), Hispano-Arab breed (Plot B) and Spanish-Sport horse breed (Plot C) by decade of birth of the ancestors. Contributions of the Spanish Arab (□), Thoroughbred (Δ), Spanish Purebred (○), Spanish Anglo-Arab (■), Hispano-Arab (●), Spanish-Sport horse (▲) and from non-listed breeds (x) ancestors are represented.

Inbreeding, average relatedness and effective population sizes

Inbreeding values of the individuals included in the reference populations were computed using both the breed exclusive and the completed pedigrees. Mean inbreeding values for the breed exclusive pedigrees were 1.09% for dAA, 0.64% for dHA, and 0.30% for dSSH whilst these figures were, respectively, 1.41%, 1.91% and 0.66% for the completed pedigrees. The mean AR coefficients, characterising the genetic identity of the individuals included in the reference populations of the dAA, dHA and dSSH breeds were 0.38%, 0.22% and 0.40% for the breed exclusive pedigrees and 1.21%, 4.62% and 0.41% for the completed pedigrees.

Table 5 gives the effective populations sizes computed from regression of the individual inbreeding coefficient over the equivalent to discrete generations in the reference population both in breed exclusive and completed datasets. The N_e values computed from regression coefficients were higher for the completed datasets in dAA and dSSH (109.5 vs. 64.5 and 135.5 vs. 47.0, respectively). N_e computed from the dHA breed exclusive dataset was 16.2 whilst that computed from the complete dataset had a meaningless negative value due to a decrease in the inbreeding coefficient in the last generation.

After checking that there was no significant regression of F over the year of birth for the animals included in the reference population to ensure that there was no trend in inbreeding, the equivalent to closed-population effective size ($_{eq}N_e$) was computed using the breed exclusive pedigrees, in order to compare the results to closed populations in terms of breeding policy. These values are shown in table 5 jointly with the migration rates in each breed. The N_e values obtained from regression coefficients in completed datasets were higher than those computed for $_{eq}N_e$. After accounting for migration rates, the computed $_{eq}N_e$ values were 39.2 for dAA, 56.3 for dHA, and 114.1 for dSSH.

Table 5. Effective population size computed from regression coefficient, migration rate (m) and equivalent to closed-population effective size ($_{eq}N_e$) computed for the reference populations (animals born in the last 12 years for the dAA breed, 10 years for dHA, and 11 years for dSSH) using the breed exclusive and the completed pedigrees.

	Spanish Anglo-Arab		Hispano-Arab		Spanish-Sport Horse	
	Breed exclusive dataset ^a	Completed dataset ^b	Breed exclusive dataset ^a	Completed dataset ^b	Breed exclusive dataset ^a	Completed dataset ^b
N_e	64.5	109.5	16.2	-	47.0	135.5
m	0.58		0.68		0.72	
$_{eq}N_e$	39.2		56.3		114.1	

^ausing breed exclusive dataset, ^busing completed dataset after including information from the parental studbooks

Between-populations differentiation

Table 6 gives the Nei's minimum distance (D_m) values and pairwise coancestry coefficients (f_{ij}) between the animals born in the last decade of the six analysed breeds: SA, SPB, TB, dAA, dHA and dSSH. The highest differentiation was

assessed between the two parental breeds with deeper pedigrees (SA and SPB; $D_m = 0.071$) whilst the lowest was between the pair TB-dSSH (0.004). With respect to the derived breeds the SA breed showed distance values varying from 0.020 (SA-dHA) to 0.035 (SA-dSSH). The coancestry assessed between SA and the derived breeds varied between 0.005 for the pair SA-dSSH and 0.032 for the pair SA-dHA. These values are in the range of the coancestry between SPB and each of dHA and dSSH (0.028 and 0.010, respectively) but are higher than those computed between TB and each of dAA and dSSH (0.001 and 0.000, respectively).

Table 6. Within-breeds coancestry (on diagonal), between-breeds coancestry matrix (above diagonal) and between-breeds Nei's minimum distance matrix (below diagonal) computed using the whole available pedigrees from the individuals born in the last decade in the Spanish Purebred, Spanish Arab, Spanish Thoroughbred, Spanish Anglo-Arab, Hispano-Arab, and Spanish-Sport horse.

	SA	SPB	TB	dAA	dHA	dSSH
Spanish Arab (SA)	0.065	0.000	0.000	0.023	0.032	0.005
Spanish Purebred (SPB)	0.071	0.076	0.000	0.000	0.028	0.010
Thoroughbred (TB)	0.041	0.035	0.005	0.001	0.000	0.000
Spanish Anglo-Arab (dAA)	0.021	0.038	0.007	0.010	0.010	0.002
Hispano-Arab (dHA)	0.020	0.018	0.016	0.009	0.027	0.006
Spanish-Sport horse (dSSH)	0.035	0.024	0.004	0.005	0.009	0.003

Discussion

Genealogical analyses in open populations have major shortcomings that make it difficult to assess the genetic variability of the stock and make inferences useful for the management of the population. Here we apply different approaches to deal with populations involving migration and to assess the genetic composition of three SA-derived Spanish horse breeds. Overall, the analyses carried out here clearly point that the information gathered in the breed exclusive pedigrees is not useful to describe the genetic scenarios of the derived populations. Differences between using the breed exclusive pedigree or the completed genealogy are a realisation of the problems of accounting for more or less deep genealogies. Overall, these figures illustrate how many between-individuals genetic relationships are lost in pedigrees without information on parental breeds. This is true for the dSSH breed regardless of whether all the parental genealogies are available but is especially true for the dHA breed in which a dramatic reduction of the genetic stock can be assessed when the full genealogies of the registered individuals are considered.

One can think such a scenario can be approached using techniques aimed at inferring pedigree information from incomplete pedigrees. There are methods that try to solve this problem based on the uncertain-parentage-matrix (UPM, Pérez-Enciso and Fernando, 1992) that considers a probabilistic correction for unknown ancestors. This method has been used very recently in Trakehner horse but no major differences were found when applied to genealogies with missing ancestries (Teegen *et al.*, 2008). Additionally, approaches to recover the inbreeding caused by incomplete pedigrees has been developed by VanRaden (1992) that assumes

that animals with missing parents have inbreeding coefficients equal to the mean of the inbreeding coefficients for animals with known parents born during the same year. Recently, Aguilar and Misztal (2008) have proposed a recursive algorithm to reduce the computing time of VanRaden's method. This approach seemed to work well only when the missing dam information was not more than 10-20% (Lutaaya *et al.*, 1999). Therefore, such techniques should not be advantageous in scenarios of open populations but in closed pedigrees with no efficient genealogies recording. Moreover, such approaches do not solve the problem of accurate assessment of the genetic contributions from a parental breed to those derived breeds.

Additionally, the possibilities of using the new information obtained in the present analysis to characterise the breeding policies of the Anglo-Arab, Hispano-Arab and Spanish Sport horse breeds can be a point of discussion. Historically, the Arab horse has been used to improve the conformation and performance characteristics of other horse populations (Bowling and Ruvinsky, 2000a) contributing to a large extent to the formation of breeds such as the Thoroughbred (Bowling and Ruvinsky, 2000b) or the Lipizzan (Zechner *et al.*, 2002). The Spanish breeds derived from the Arab horse are: dAA, dHA and dSSH. The introgression of SA individuals into these derived breeds is significant and, in most cases, balanced between stallions and mares (see Table 4). The use of reproductive SA individuals in the derived breeds has been continuous up to the present day and, in fact, the derived breeds can not be considered as separated genetic entities. Major SA founders and ancestors identified for the derived breeds (especially for dAA and dHA) were born recently between 1975 and 1994 and no unbalanced use of reproductive individuals belonging to the derived breeds was detected until recently (see Figures 2 and 3).

Completeness and gene origin

Any attempt to compare the parameters characterising gene origin of the analysed breeds with the results available in the literature is not straightforward because the latter are computed on closed horse populations with deep studbooks (see Valera *et al.*, 2005 and Zechner *et al.*, 2002 for reviews). Also, this kind of comparison must be interpreted with caution since these parameters depend on the time from foundation of the Studbook and the quality of pedigree data recording. Recently, Cervantes *et al.* (2008a) reported values of total number of founders, effective number of founders and effective number of ancestors in the SA breed of, respectively, 860, 39.5 and 13. The SA breed is known to have a deep studbook founded in the 19th century that remains open for the inclusion of Arab individuals from foreign countries. However, the figures corresponding to the breeds analysed here are higher than those reported for the SA breed (Cervantes *et al.*, 2008a) with the rough exception of the completed dataset of the dHA breed that have a considerably higher pedigree depth than the others. Differences found for the dSSH breed can be explained by the recent foundation of the studbook and the lower quality of the available pedigree data. However, differences between dAA and dHA are mainly due to the higher quality of the genealogies available for the SPB ancestors of dHA individuals.

Inbreeding, Average Relatedness and effective population size

Both the average inbreeding and AR computed for the reference populations defined in the pedigrees analysed here are highly affected by the degree of completeness of the genealogies. Some of the individuals that appeared as a founder in the breed exclusive dataset are heavily inbred. In open populations it seems to be not necessary the control in the breeding policy; a non inbred offspring will be obtained from two unrelated highly inbred parents. However mating among progeny in later generations from related founder animals may increase inbreeding in very short time and reduce effective population size. Mean F computed using the completed datasets is from 30% (dAA) to 300% (dHA) higher than that assessed using the breed exclusive datasets. The mean AR s (which is roughly twice the expected average F in the case of random mating, Goyache *et al.*, 2003; Gutiérrez *et al.*, 2003; Gutiérrez and Goyache, 2005) increase dramatically in the breeds with deeper pedigrees (dAA and dHA), thus showing that breeders plan matings in these breeds using related parental individuals, especially in dHA, because of the between-individuals relationship existing in the SPB studbook (Valera *et al.*, 2005). This is probably happening because breeders ignore that animals from parental breeds are highly related since they do not have access to the complete dataset.

The N_e computed with the breed exclusive pedigrees had values of 64.5, 16.2 and 47.0 for, respectively, the dAA, dHA and dSSH breeds (Table 5). However, these figures do not address the 'real' N_e because they ignore the relationships between individuals appearing as founders, as well as the breeding policy carried out in the parental breeds. These effective population sizes were computed using a regression approach that has been shown to fit properly with real pedigrees with a stable increase in inbreeding (Gutiérrez *et al.*, 2003). However, in the scenario analysed here the computed values are highly affected by migration and mating policies. This is especially true in the case of short pedigrees in which a small number of individuals with more complete pedigrees are highly inbred, such as in the case of the breed exclusive dataset of the dHA breed which includes only 84 inbred individuals with mean F of 13.7%. Moreover, the regression approach assumes that the reference population is the last and more inbred generation. This is not realistic in the case of migration as highlighted when the completed datasets were analysed. The N_e 's obtained using regression coefficients with the completed datasets increase both in the dAA and the dSSH breeds. This fact shows that the increase in inbreeding in the parental breeds is lower than in the derived breeds, leading to a global lower increase in inbreeding in the completed datasets. Furthermore, it must be noted that somewhere in the pedigree of each of the individuals of the completed pedigrees a mating between animals from a different parental breed could occur with an intermediate result of null inbreeding. An example of this fact is observed in the meaningless N_e value computed for dHA when the completed dataset was considered. This is because the parental generation was more inbred than the reference population (6.38% vs 1.91%).

In order to compare the results obtained here with closed populations similar in size and breeding policy (55.4±4.96 for SPB and 39.1±4.19 for SA; Cervantes *et*

al., 2008b), N_e should be computed taking into account migration rates (Falconer and Mackay, 1996). The $_{eq}N_e$'s given in Table 5 are consistent with the history of the breeds and are from 36% (dAA) to 84% (dSSH) of their respective values computed by regression in completed datasets.

Between-populations differentiation

The between-breeds differentiation assessed via D_m is highly affected by the within-breeds coancestry. Previously unknown ancestors in derived populations with shallower pedigrees are not expected to increase the between-populations coancestry since different breeds do not share common ancestors. However, these ancestors would increase the coancestry within populations. In other words, between-breeds distances are underestimated when one of them has poor pedigree knowledge. Differentiation between the SA and the Arab-derived breeds analysed here is not negligible because the average within-breeds genetic coancestry exceeds the between-breeds coancestry. This can be also observed in the case of the SPB-derived breeds. However, the very low within-breed coancestry assessed in TBs (0.005) gives lower D_m values than expected. In this respect, the within-breeds coancestry values and, especially, the between-breeds coancestry matrix are more informative on the mating policies and histories of the breeds. Note that coancestries are half the additive genetic relationships that are, in turn, the percentage of genes shared between two individuals or, as here, populations. Taking this into consideration dAA and dHA populations would share a higher percentage of genes (4.6% for dAA and 6.4% for dHA) with the present SA population than with, respectively, the present TB and SPB populations. These results are not consistent with the classical approach used before to assess founder genetic contributions to a population thus showing that the gene origins of dAA and dHA are closer to, respectively, the TB and the SPB breeds than to the SA breed (see Table 4). An alternative explanation is that the TB and SPB individuals used for reproduction in the dAA and dHA breeds are not the same as those used for reproduction within the parental breeds, particularly in the case of TBs. In any case, this analysis clearly shows that estimates of genetic contribution from parental to derived breeds are biased in the case of open studbooks. In this respect, only when studbook information from parental breeds is not available, contrary to what happened in this study, the performance of such estimations could be recommended.

Conclusions

We carried out pedigree analyses in three Spanish Arab-derived horse breeds. The methodologies used in this work illustrate the difficulties in carrying out genealogical studies in open populations.

The analysis carried out with the exclusive datasets was not enough to characterise the genetic variability of derived breeds, but the analysis made by completed datasets attained better results. The available genetic variability in the SA-derived breeds was poorer than expected in populations subject to continuous migration due to the use for reproduction of a limited number of families from the parental horse breeds. This is reflected by the major decrease assessed for

parameters such as effective number of founders, effective number of ancestors and founder genome equivalents when the completed pedigrees are considered. This is particularly true in the dHA breed which is considered an endangered breed in Spain. From a breeding perspective, the present study emphasizes the need to include the complete genealogies of the individuals from parental breeds used for reproduction in open populations to avoid unobserved loss of genetic variability putting at risk the genetic stocks.

The need to account for migration rates to compute N_e is empirically supported here because the regression approaches tested to compute N_e did not fit populations under migration. Moreover, we point out that ascertaining the genetic contributions from a population to another via founder analyses can not be consistent with between-populations coancestries and can therefore be questioned. As for between-studbooks connectedness, it has been shown that further studies are needed to assess possible joint evaluation for common traits included in the corresponding selection schemes. The essayed methodologies were shown to be useful to deal with animal populations under migration. Misleading conclusions can be deduced if the whole relationship among migrants is unknown.

Acknowledgments

The authors wish to thank the ‘Subdirección General de Medios de Producción Ganaderos’ of the Spanish Ministry of Agriculture for their support. We also thank the ‘Fondo de Explotación de los Servicios de Cría Caballar y Remonta’ of the Spanish Ministry of Defence and the Spanish Arab, Spanish Purebred and Spanish Sport Horse Breeders’ Association for their collaboration. This work was partially funded by a grant from the Junta de Andalucía given to I. Cervantes (BOJA, nº 120, 21/06/2004).

References

- Aguilar, I., Misztal, I. (2008) Technical Note: Recursive Algorithm for Inbreeding Coefficients Assuming Nonzero Inbreeding of Unknown Parents. *J. Dairy Sci.*, **91**, 1669–1672.
- Azor, P.J., Valera, M., Gómez, M.D., Goyache, F., Molina, A. (2007) Genetic characterization of the Spanish Trotter horse breed using microsatellite markers. *Genet. & Mol Bio.*, **30** (1), 37-42.
- Ballou J.D., Lacy, R.C. (1995) Identifying genetically important individuals for management of genetic variation in pedigreed populations. In: Ballou, J.D., Gilpin, M., Foose, T.J. (ed). Population management for survival and recovery: analytical methods and strategies in small population management. New York: Columbia University Press, pp 76-111.
- Bowling, A.T., Ruvinsky, A. (2000a) The genetic resources and their conservation. In. “The Genetic of the horse”. CAB International. UK, pp 387-410.
- Bowling, A.T., Ruvinsky, A. (2000b). Genetic Aspects of Domestication. In. “The Genetic of the horse”. CAB International. UK, pp 25-52.
- Boichard, D., Maignel, L., Verrier, E. (1997) The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Sel. Evol.*, **29**, 5-23.

- Caballero, A., Toro, M.A. (2000) Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genet. Res. Camb.*, **75**, 331-343.
- Caballero, A., Toro, M.A. (2002) Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv. Genet.*, **3**, 289-299.
- Cervantes, I., Molina, A., Goyache, F., Gutiérrez, J.P., Valera, M. (2008a) Population history and genetic variability in the Spanish Arab Horse assessed via pedigree analysis. *Livest. Sci.*, **113**, 24-33.
- Cervantes, I., Goyache, F., Molina, A., Valera, M., Gutiérrez, J.P. (2008b) Application of individual increase in inbreeding to estimate realised effective sizes from real pedigrees. *J. Anim. Breed. Genet.*, **125**, 301-310.
- Falconer, D.S., Mackay, F.C. (1996) Introduction to Quantitative Genetics. Longman Group Ltd. England.
- Goyache, F., Gutiérrez, J.P., Fernández, I., Gómez, J., Álvarez, I., Diez, J., Royo, L.J. (2003) Using pedigree information to monitor genetic variability of endangered population: the Xalda sheep breed of Asturias as an example. *J. Anim. Breed. Genet.*, **120**, 95-103.
- Gutiérrez, J.P., Altarriba, J., Díaz, C., Quintanilla, R., Cañón, J., Piedrafita, J. (2003) Genetic analysis of eight Spanish beef cattle breeds. *Genet. Sel. Evol.* **35**, 43-64.
- Gutiérrez, J.P., Goyache, F. (2005) A note on ENDOG: a computer program for analysing pedigree information. *J. Anim. Breed. Genet.* **122**, 172-176.
- Gutiérrez, J.P., Marmí, J., Goyache, F., Jordana, J. (2005) Pedigree information reveals moderate to high levels of inbreeding and a population genetic structure in the Catalanian donkey breed. *J. Anim. Breed. Genet.*, **122**, 378-386.
- Hamann, H., Distl, O. (2008) Genetic variability in Hanoverian warmblood horses using pedigree analysis. *J. Anim. Sci.*, **86**, 1503-1513.
- James, J. (1972) Computation of genetic contributions from pedigrees. *Theor. Appl. Genet.*, **42**: 272-273.
- Lacy, R.C. (1989) Analysis of Founder Representation in Pedigrees: Founder Equivalent and Founder Genome Equivalents. *Zoo. Biol.*, **8**, 111-123.
- Lutaaya, B. E., Misztal, I., Bertrand, J. K, Mabry, J. W. (1999) Inbreeding in populations with incomplete pedigrees. *J. Anim. Breed. Genet.*, **116**, 475-480.
- Malécot, G. (1948) Les Mathématiques de l'Hérédité. Masson et Cie, Paris.
- Maxwell, J. (1995) Spanish Arabian Horse Families. Alexander Heriot & Co. Ltd, England.
- MacCluer, J., Boyce, B., Buke, L., Weitzkamp, D., Pfenning, A., Parsons, C. (1983) Inbreeding and pedigree structure in Standardbred horses. *J. Heredity* , **74**, 394-399.
- MAPyA. (2003) Estudio y caracterización genética del sector equino en España. Ministerio de Agricultura Pesca y Alimentación.
(<http://www.mapa.es/app/Equino/Informacion/Infsector.aspx?lng=e>). España.

- Moureaux, S., Verrier, E., Ricard, A., Meriaux, J.C. (1996) Genetic variability within French race and riding horse breeds from genealogical data and blood marker polymorphisms. *Genet. Sel. Evol.*, **28**, 83-102.
- Pérez-Enciso, M. (1995) Use of the uncertain relationship matrix to compute effective population size. *J. Anim. Breed. Genet.*, **112**, 327-332.
- Steen, A.K. (2006) A glorious task. A tale of the breed. Escandón Impresores. Seville. Spain.
- Teegen, R., Edel, C., Thaller, G. (2008) Population structure of the Trakehner Horse breed. *Animal*, doi:10.1017/S1751731108003273.
- Valera, M., Molina, A., Gutiérrez, J.P., Gómez, J., Goyache, F. (2005) Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain. *Livest. Prod. Sci.*, **95**, 57-66.
- VanRaden, P. M. (1992) Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *J. Dairy Sci.*, **75**, 3136-3144.
- Wright, S. (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97-159.
- Zechner, P., Sölkner, J., Bodo, I., Druml, T., Baumung, R., Achmann, R., Marti, E., Habe, F., Brem, G. (2002) Analysis Diversity and population structure in the Lipizzan horse bred based on pedigree information. *Livest. Prod. Sci.*, **77**, 137-146.

4.3. ARTÍCULO 3: INDIVIDUAL INCREASE IN INBREEDING ALLOWS
ESTIMATING EFFECTIVE SIZES FROM PEDIGREES

Genetic Selection Evolution (2008), 40: 359-378

J. P. Gutiérrez^{1*}, I. Cervantes¹, A. Molina², M. Valera³ and F. Goyache⁴

¹Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro s/n, E-28040-Madrid, Spain.

²Departamento de Genética, Universidad de Córdoba, Ctra. Madrid-Cádiz, km 396^a, E-14071 Córdoba, Spain.

³Departamento de Ciencias Agro-Forestales, EUITA, Universidad de Sevilla, Ctra. Utrera km 1, E-41013 Sevilla, Spain.

⁴SERIDA-Somío, C/ Camino de los Claveles 604, E-33203 Gijón (Asturias), Spain.

(Received 14 May 2007; accepted 9 January 2008)

*Corresponding author: gutgar@vet.ucm.es

Abstract

We present here a simple approach to obtain reliable estimates of the effective population size in real world populations via the computation of the increase in inbreeding for each individual (ΔF_i) in a given population. The values of ΔF_i are computed as t -root of $1 - (1 - F_i)$ where F_i is the inbreeding coefficient and t is the equivalent complete generations for each individual. The values of ΔF computed for a pre-defined reference subset can be averaged and used to estimate effective size. A standard error of this estimate of N_e can be further computed from the standard deviation of the individual increase in inbreeding. The methodology is demonstrated by applying it to several simulated examples and to a real pedigree in which other methodologies fail when considering reference subpopulations. The main characteristics of the approach and its possible use are discussed both for predictive purposes and for analysing genealogies.

Effective size/ increase in inbreeding/ overlapped generations/ genetic contributions

1. INTRODUCTION

The effective population size (N_e), defined as ‘the size of an idealized population which would give rise to the rate of inbreeding, or the rate of change in variance of gene frequencies observed in the population under consideration’ [27], is a key parameter in conservation and population genetics because of its direct relationship with the level of inbreeding, fitness and the amount of genetic variation lost due to random genetic drift [5, 7]. As a consequence, N_e is usually considered as a useful criterion for classifying the livestock breeds according to the degree of endangerment [6, 8].

When genealogies are available, the effective population size can be estimated from the increase in inbreeding (ΔF) between two discrete generations as

in $N_e = \frac{1}{2\Delta F}$, with $\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$, where F_t and F_{t-1} are the average inbreeding

at t and $t-1$ generations [7]. The increase in inbreeding is constant for an ideal population of constant size with no migration, no mutation and no selection over discrete generations. However, in real populations with overlapping generations, the number of males and females are usually different and non-random matings are the rule, making ΔF a difficult parameter to deal with [7]. In most cases the definition of a ‘previous’ generation is quite difficult to establish. In fact, taking the average inbreeding of a pre-defined reference subpopulation and refer it to the founder population in which inbreeding is null by definition, fits poorly in any given real population and is only acceptable in small populations with shallow pedigree files [1, 10, 11] leading to the risk of overestimating the actual effective population size.

Some attempts have been proposed to overcome these challenges in the real world, namely the computation of N_e from the variances of family sizes of males and females [7, 13, 14] or the use of the regression coefficient of the individual inbreeding coefficients on the number of generations known for each animal as an estimate of ΔF [11]. In a scenario of overlapping generations, computation of N_e based on family variances unrealistically ignores population subdivision and

several other causes of variation of the parameter, such as mating between relatives, migration, or different representation of founders. Most methodologies applied to compute N_e under overlapping generations are also affected by the difficulties in fitting individuals to generations because data over time usually appear as registered by year regardless of when the renewal of the population is done at a generation interval. On the other hand, the computation of regression

coefficients with the aim of approximating $\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$, has also the difficulty of

defining the 'previous' generation with respect to the identified reference subpopulation. The estimation of effective size could be approximated by using

$1 - F_t = \left(1 - \frac{1}{2N_e}\right)^t$ to derive its value from a log regression of $(1 - F)$ over a

generation number [20], thus avoiding the need to define a previous generation. When the value of t is difficult to establish, this can be estimated by considering the year of birth as t and further correcting by the length of the generation interval [20]. However, variations in the breeding policy, such as planning matings to minimize coancestry after a period in which mating between close relatives was preferred, can lead to a temporal decrease in average inbreeding. When the animals of interest are those born in the period in which the inbreeding decreased, methods based on assessing the increase in inbreeding would lead to negative values of N_e .

Moreover, in real populations in which selection is likely to occur, increase in inbreeding is not a consequence of the sole accumulative change of gene frequency of a neutral gene over generations but of the long-term genetic contributions made by the ancestors [25, 26]. In fact, the average inbreeding coefficient of a current reference subpopulation depends on both the number of generations separating this reference subpopulation from the founder population, and how rapidly the inbreeding accumulates.

The concept of effective size can therefore be interpreted not only as a useful parameter to predict inbreeding, but also as a tool to analyse genealogies [5]. Many attempts have been made to deal with the different real world scenarios in order to obtain reliable estimates of the effective population size [4, 5]. However, there is no standard method for general application to obtain the effective population size. Here we present a straightforward approach to deal with this task by the computation of the increase in inbreeding for each individual (ΔF_i) in a given population. The values of ΔF_i are useful to obtain reliable estimates of N_e . The N_e estimated this way roughly describes the history of the pedigrees in the population of interest. The approach directly accounts for differences in pedigree knowledge and completeness at the individual level but also, indirectly, for the effects of mating policy, drift, overlap of generations, selection, migration and different contributions from different number of ancestors, as a consequence of their reflection in the pedigree of each individual in the analysed population. This approach, which is based on the computation of individual increase in inbreeding, also makes it possible to obtain confidence intervals for the estimates of N_e .

2. MATERIALS AND METHODS

2.1. Individual increase in inbreeding

We will start from a population with a size of N individuals bred under conditions of the idealized population [7]. Under these conditions the inbreeding at a hypothetical generation t can be obtained by [7]:

$$F_t = 1 - (1 - \Delta F)^t \quad (1)$$

The idea presented here is to calculate inbreeding values and a measure of equivalent discrete generations for each animal belonging to a subgroup of animals of interest (the so called reference subpopulation) in a scenario with overlapping generations. Then, from (1), and equating the individual inbreeding coefficient to that for a hypothetical population with all individuals having the same pedigree structure ($F_i = F$), an individual increase in inbreeding (ΔF_i) can be defined as:

$$\Delta F_i = 1 - \sqrt[t]{1 - F_i} \quad (2)$$

where t is the 'equivalent complete generations' [3, 18] calculated for the pedigree of the individual as the sum over all known ancestors of the term of $(\frac{1}{2})^n$, where n is the number of generations separating the individual from each known ancestor. Notice that, on average, for a given reference subpopulation, t is equivalent to the 'discrete generation equivalents' proposed by Woolliams and Mäntysaari [24], thus characterising the amount of pedigree information in data sets with overlapping generations. Parameter t has been widely used to characterise pedigree depths both in actual [1, 9, 21] and simulated datasets [2].

The set of ΔF_i values computed for a number of individuals belonging to the reference subpopulation can be used to estimate the N_e regardless of the presence of individuals which would be assigned to different discrete generations according to their pedigree depth. The ΔF_i values of the individuals belonging to the reference population can be averaged to give $\overline{\Delta F}$. From this, a mean effective population

size $\overline{N_e}$ can be straightforwardly computed as $\overline{N_e} = \frac{1}{2\overline{\Delta F}}$. Notice that this way of

computing effective population size is not dependent on the whole reference subpopulation mating policy but on the matings carried out throughout the pedigree of each individual.

Moreover, since we are assuming a different individual increase in inbreeding for each individual i in the reference subpopulation, ascertaining the confidence on the estimate of $\overline{\Delta F}$ is also feasible, and the corresponding standard error can be easily computed. Kempen and Vliet [17] described how the variance of the ratio of the mean of two variables x and y can be approximated using a Taylor series expansion. Assigning in our case $x = 1$, and $y = 2\Delta F$, we can obtain the standard

error of $\overline{N_e}$ as $\sigma_{\overline{N_e}} = \frac{2}{\sqrt{N}} \overline{N_e}^2 \sigma_{\Delta F}$, with N being the number of individuals in the reference subpopulation, $\sigma_{\Delta F}$ the standard deviation of ΔF and $\sigma_{\overline{N_e}}$ the standard error of $\overline{N_e}$. It can also be easily shown that this is equivalent to assuming that N_e has the same coefficient of variation as ΔF .

2.2. Other methods to estimate N_e using pedigree information

Various additional approaches have been used to compare estimates of N_e obtained from individual increase in inbreeding. First, N_e was estimated from the rate of inbreeding (ΔF) or the rate of coancestry (Δf) observed between two discrete generations as, respectively, $N_e = \frac{1}{2\Delta F}$ and $N_e = \frac{1}{2\Delta f}$, with

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}} \text{ and } \Delta f = \frac{f_t - f_{t-1}}{1 - f_{t-1}}, \text{ where } F_t \text{ and } F_{t-1} \text{ and } f_t \text{ and } f_{t-1} \text{ are the}$$

average inbreeding and the average coancestry at the t and $t-1$ generations. Moreover, N_e was estimated from the variances of family sizes as [13]:

$$\begin{aligned} \frac{1}{N_e} = \frac{1}{16ML} & \left[2 + \sigma_{mm}^2 + 2 \left(\frac{M}{F} \right) \text{cov}(mm, mf) + \left(\frac{M}{F} \right)^2 \sigma_{mf}^2 \right] \\ & + \frac{1}{16FL} \left[2 + \left(\frac{F}{M} \right)^2 \sigma_{fm}^2 + 2 \left(\frac{F}{M} \right) \text{cov}(fm, ff) + \sigma_{ff}^2 \right] \end{aligned} \quad (3)$$

where M and F are the number of male and female individuals born or sampled for breeding each time period, L the average generation interval σ_{mm}^2 and σ_{mf}^2 are the variances of the male and female offspring of a male, σ_{fm}^2 and σ_{ff}^2 are the variances of the male and female offspring of a female, and $\text{cov}(mm, mf)$ and $\text{cov}(fm, ff)$ the respective covariances. Note that the family size of a parent (male or female) consists of its number of sons and daughters kept for reproduction [14]. The three approaches described above were applied to the simulated pedigree files with the data structured in discrete generations.

When datasets with no discrete generations were analysed, N_e was estimated from the variances of family sizes but also from ΔF using three different approaches: first, following Gutiérrez et al. [11], the increase in inbreeding between two generations ($F_t - F_{t-1}$) was obtained from the regression coefficient (b) of the average inbreeding over the year of birth obtained in the reference subpopulation, and considering the average generation interval (l) as follows:

$$F_t - F_{t-1} = l \times b$$

with F_{t-1} computed from the mean inbreeding in the reference subpopulation (F_t) as:

$$F_{t-1} = F_t - l \times b$$

Second, in a similar way N_e was obtained using directly t instead of considering the generations through generation intervals. Using this approach, N_e was computed from the regression coefficient (b) of the individual inbreeding values over the individual equivalent complete generations approximating t . In this case

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}} \approx \frac{b}{1 - (F_t - b)} \quad (4)$$

with F_t being the average F of the reference subpopulation.

Finally, we applied the approach by Pérez-Enciso [20] to estimate N_e via a log

regression of $(1 - F)$ (obtained from (1) as $1 - F_t = \left(1 - \frac{1}{2N_e}\right)^t$) on generation

number. When datasets with no discrete generations were analysed N_e was estimated by a log regression of $(1 - F)$ on the date of birth and then divided by the generation interval [20].

2.3. Examples

The methodology is demonstrated by applying it to four simulated examples embracing a wide range of typical theoretical scenarios.

The simulated datasets evolved during 50 generations (200 periods of time in the third example under overlapping) from a founder population consisting of 200 individuals under the following assumptions:

- i) The founder population is formed by the same number of individuals of two different sexes. A total of 100 males and 100 females are born in each generation and act as parents of the following generation under random mating with the individuals of the other sex and no differential viability or fertility. The theoretical N_e excluding self-fertilization is the number of individuals + $\frac{1}{2}$ (200.5) [7].
- ii) Like simulated population i) but splitting the populations in 4 different subpopulations consisting of 25 males and 25 females evolving separately after generation 25. The theoretical N_e is as i) before subdivision. After that, the theoretical N_e for each subpopulation is 50.5
- iii) Like simulated population i) but limiting the renewal of reproductive individuals to 25 males and 25 females each period of time and allowing the reproductive individuals to have offspring during 4 consecutive periods. Under overlapping

generations the expected N_e can be derived from the expression

$$N_e \approx \frac{8N}{V_{km} + V_{kh} + 4} L \quad [7, 13] \quad \text{where } N_C \text{ is the number of}$$

reproductive individuals included in the reference subpopulation (50), V_{km} and V_{kf} are, respectively, the variances of family sizes of reproductive males and females ($V_{km} = V_{kf} = 2$ under random conditions), and L is the generation length in units of the specified time-interval (2.5). Here N_e equals to 125.

- iv) Like simulated population i) but all parents having two offspring in the next generation. This is a case where mating is random but the variance of family sizes does not follow a Poisson distribution. The expected value of N_e computed from the

$$\text{expression: } N_e \approx \frac{8N}{V_{km} + V_{kh} + 4} L \quad [7, 13], \text{ after equaling } V_{km} =$$

$V_{kh} = 0$, is 400.

The simulated pedigree files i) to iv) listed above are expected to characterise classical theoretical scenarios of populations evolving randomly with two sexes (i), population subdivision (ii), overlapping generations (iii), and non-Poisson variance of family sizes (iv). Within each pedigree file a reference subset was defined as the last 400 animals born.

Additionally, the pedigree file of the Carthusian strain of the Spanish Purebred horse was used to demonstrate the methodology on a real example. It is a subpopulation of the pedigree file of Andalusian horse (SPB, Spanish Purebred horse) [22] and included a total of 6,318 individuals since the foundation of the studbook. This population is expanding with 45 % of the registered individuals born in the last twenty years. This period of time is roughly the last two generations (Fig. 1) [22]. The pedigree knowledge is reasonably high: 95% of ancestors tracing back seven generations were known and the mean equivalent complete generations for the animals born in the last decade was 9.1.

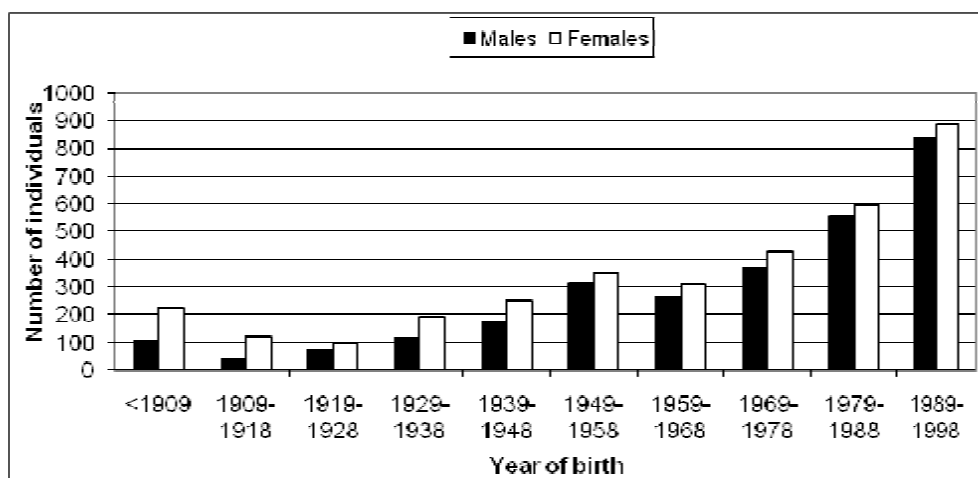


Figure 1. Evolution of registered individuals per year of birth in the Carthusian subpopulation.

The Carthusian strain was chosen as a real example of inbred population, because it had been subjected to a planned mating strategy using the minimum coancestry approach starting from the 1980's [22]. Due to this mating policy a decrease in the mean inbreeding coefficients along the period involving the last generation was also found [22]. This enables testing for the possible influence of a particular supervised breeding policy on N_e . Two Reference Subset (RS) were defined in the Carthusian pedigree file: the individuals born in the last 10 years of available records (RS_{10}), and the individuals born in a given period of years allowing their use for reproduction (1977-1989; RS_{77-89}). The pedigree files of the fitted RS's were also edited to include only individuals with 4 equivalent generations or more, and 8 equivalent generations or more. The main parameters describing the Carthusian pedigree file are given in Table I.

Table I. Number of individuals (N), average number of equivalent generations and standard deviation ($t \pm s.d.$), maximum number of equivalent generations (Max. t), average inbreeding (F ; in percent), number of male and female reproductive individuals and average family size for males and females (in brackets), and variances of family sizes for reproductive males (V_m) and females (V_f) for the whole Carthusian pedigree file (WP) and their reference (RS) subset used as example in the present analyses.

	N	$t \pm s.d.$	Max. t	F (%)	Stallions	Mares	V_m	V_f
WP	6,318	6.6 \pm 2.75	10.9	13.0	424 (4.9)	933(2.3)	46.96	2.85
RS_{10}	1,721	9.1 \pm 0.68	10.9	18.6	1 ^a (1.0)	5 ^b (1.2)	0	0.2
RS_{77-89}	1,464	8.2 \pm 0.64	9.8	17.5	97 ^a (4.1)	16 ^b (2.1)	21.98	1.71

RS_{10} : animals born in the last decade. RS_{77-89} : animals born between the years 1977 and 1989.

^aindividuals born in the defined period that acted sequentially as stallions.

^bindividuals born in the defined period that acted sequentially as mares.

2.4. Program used

The analyses were performed using the ENDOG program (current version v4.4) [12] which can be freely downloaded from the World Wide Web at http://www.ucm.es/info/prodanim/html/JP_Web.htm.

3. RESULTS

The results from the analyses carried out on the simulated pedigree files are summarised in Figure 2. A discontinuous line was drawn for the theoretical effective size as reference under the different scenarios. In the case of subdivision (ii) the theoretical effective population size was also computed as the harmonic mean over generations which expresses the expected N_e under descriptive rather than predictive purposes.

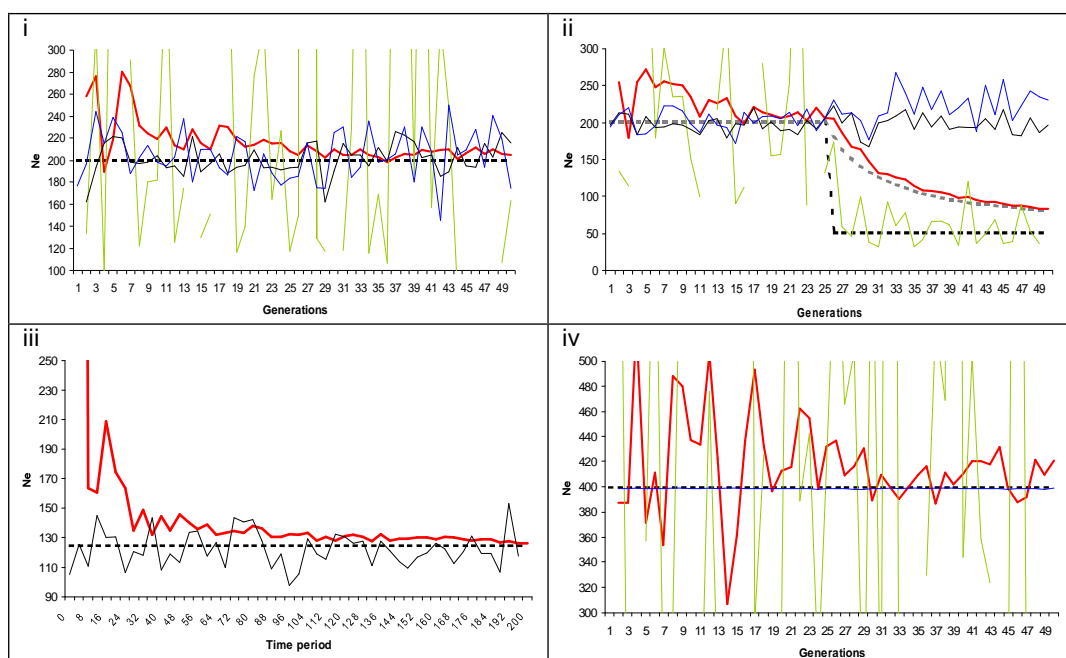


Figure 2. Variation over the time of the estimates of N_e in four simulated populations. i) ideal population assuming two sexes; ii) population subdivision; iii) overlapping generations; and iv) non Poisson variance of family sizes.

---Theoretical N_e , --- Theoretical N_e by harmonic mean, — N_e from rate of inbreeding (ΔF), — N_e from the rate of increase in coancestry (Δf), — N_e from the variance in the family sizes, — N_e from individual increase in inbreeding.

Note the erratic behaviour over generations, in Figure 2, of the N_e computed using the rate of inbreeding in the idealized (plot i) and non-Poisson offspring size variance (plot iv) populations. N_e tended to fit better in the case of population subdivision (plot ii) and could not be used under a scenario with overlapping generations (plot iii). This erratic behaviour was caused by the use of a single

replicate in the simulation and could be overcome by using the harmonic mean of N_e by generations. Estimations of N_e based on an increase in coancestry, are, however, more precise because they are computed using much more data (all pairs of individuals rather than the number of individuals), and is almost exact in the case of all animals having identical offspring size. N_e values computed using Δf and those based on variance of family size, tended to fit well in the idealized population and in the case of overlapping generations, but it failed when considering the case of population subdivision because the method ignores that such a subdivision exists. After about eight generations, performance of the individual increase in inbreeding tended to fit better than those based on Δf and variance of family sizes in the idealized population. In the case of a population subdivision the N_e computed from individual increase in inbreeding fits very closely to the N_e computed as the harmonic mean of the number of animals over generations for descriptive purposes and N_e using rate of inbreeding tended to approximate the theoretical N_e for predictive purposes. The computed effective population size using ΔF_i accounts for all historical pedigree of the individuals and the obtained N_e summarises all the genealogical information of each individual. Therefore, the genealogies recorded before subdivision weighs much more at the time closer to the population fission but its weight decreases with the accumulation of generations. If the estimation of N_e from the generations after fission is carried out for predictive purposes the harmonic mean of N_e throughout generations would be preferred rather than N_e based on individual increase in inbreeding since this converges much slower towards the 'theoretical' N_e . However, the latter one addresses better the history of the population if the estimation of N_e is carried out for descriptive purposes. In the case of overlapping generations (plot iii) $\overline{N_e}$ had a better performance than that based on variance of family size, that resulted in a more variable N_e . In the non-Poisson case in which all individuals have an offspring size of 2 (plot iv), all the involved methodologies tended to give the correct value. Regarding dispersion, the estimates of N_e based in individual increase in inbreeding were intermediate between those from the rate of inbreeding (ΔF) and from the rate of coancestry (Δf). Obviously the method based on variance of family size was, in this case, exact.

Table I gives the main parameters describing the real pedigree file. The pedigree size was 6,318 and the size for the fitted RS's was 1,464 individuals for RS₇₇₋₈₉ and 1,721 for RS₁₀. The mean equivalent generation \pm standard deviation was: 6.6 ± 2.75 for the whole pedigree (WP), 9.1 ± 0.68 for RS₁₀ and 8.2 ± 0.64 for RS₇₇₋₈₉.

Figure 3 shows the evolution of the mean inbreeding, mean individual increase in inbreeding, mean equivalent generations, and $\overline{N_e}$ across years of birth for the whole Carthusian pedigree file. It can be noted that mean inbreeding became approximately stable in the last generation interval, whilst mean equivalent generations increased leading to a reduction in the mean ΔF_i during this period. The flat or slightly negative trend of inbreeding coefficients would lead to illogical estimates of N_e when using methods based on regression of inbreeding on either generations or year of birth. However, N_e obtained from individual increase in

inbreeding remains approximately stable since pedigree knowledge achieves about five equivalent generations.

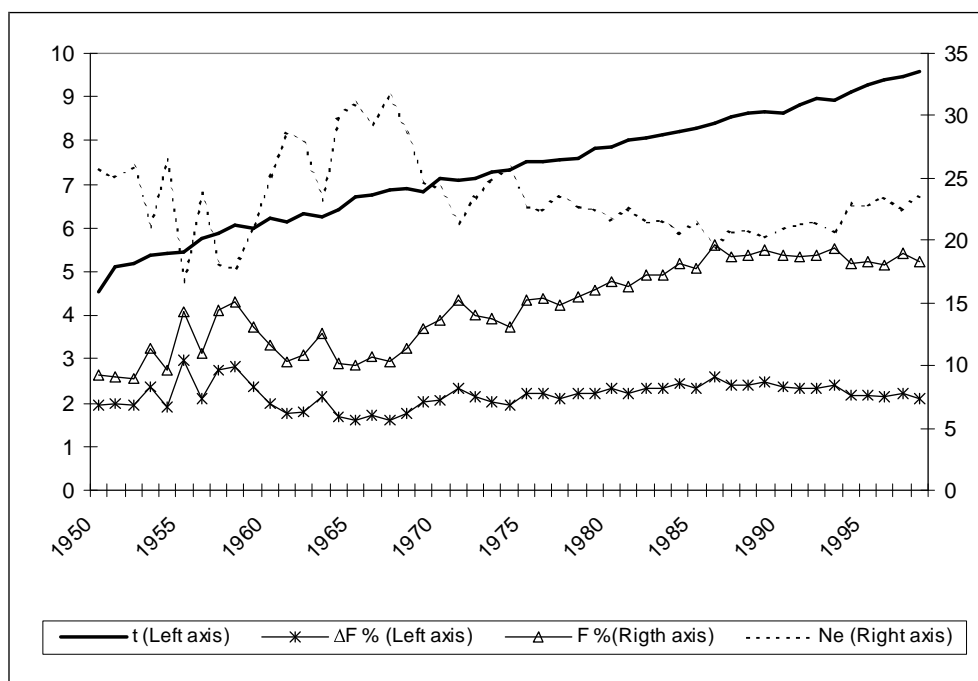


Figure 3. Evolution of the mean inbreeding, mean individual increase in inbreeding, mean equivalent generations, and $\overline{N_e}$ across year of birth for the whole Carthusian pedigree file.

Table II gives the estimates of N_e obtained using regression of the individual coefficients of inbreeding on equivalent generations, variance of family sizes and individual increase in inbreeding (ΔF) in the whole pedigree file and the defined RS's of the Carthusian horse. The estimates obtained using variances in family sizes are quite larger than those obtained using both regression on equivalent generations and ΔF .

The values for $\overline{N_e}$ were stable regardless of any restrictions on the pedigree depth of the individuals included in the pedigree file or the corresponding RS's, whilst the N_e obtained from regression of F on equivalent generations showed a non consistent behavior. It is possible to find a noticeable increase of the N_e (40.0) when RS₁₀ includes individuals with 4 or more equivalent generations in the pedigree file but also negative values of N_e (-142.2) when RS₁₀ includes individuals with 8 or more equivalent generations in the pedigree file. Note again that a negative estimation of N_e can be obtained when the increase in inbreeding is obtained by regression of the inbreeding coefficient on the year of birth and younger individuals are less inbred than the older. Thus, the N_e obtained depends partially, on the effect of the changes in the mating policy. In the Carthusian

population, the criterion of minimal coancestry has recently been used to define the mating policy of this population [22]. However, the N_e obtained by individual increase in inbreeding shows a stable value of about 22.

Table II. Estimates of N_e using the whole pedigree file (WP) and reference subsets fitted using the individuals born in the last 10 years (RS₁₀) or in a given period of years (RS₇₇₋₈₉) obtained using regression of the individual coefficients of inbreeding on equivalent generations [$b(t)$], variance of family sizes (V) and individual increase in inbreeding ($\overline{N_e}$). Analyses were carried out without pedigree depth restrictions, and using only individuals with 4 equivalent generations (t) or more and 8 equivalent generations (t) or more.

Population	No pedigree depth restriction			$t \geq 4$			$t \geq 8$		
	$N_e [b(t)]$	$N_e (V)$	$\overline{N_e} \pm \text{s.d.}$	$N_e [b(t)]$	$\overline{N_e} \pm \text{s.d.}$	N	$N_e [b(t)]$	$\overline{N_e} \pm \text{s.d.}$	N
WP	19.5	35.9	26.4±0.26	20.4	22.4±0.17	5,172	273.5	21.4±0.14	2,518
RS ₁₀	33.4		22.2±0.18	40.0	22.2±0.18	1,719	-142.2	22.0±0.18	1,671
RS ₇₇₋₈₉	12.2	88.1	21.4±0.21	11.6	21.3±0.21	1,463	21.9	20.3±0.22	924

RS₁₀: animals born in the last decade. RS₇₇₋₈₉: animals born between the years 1977 and 1989.

The overall behaviour of $\overline{N_e}$ is summarized in Table III. In this table a reference subset as the last 400 individuals born was defined, which involves the last two generations in the simulated examples under discrete generations. Together with the simulated examples the real whole pedigree file and the RS's defined in the Carthusian pedigree file were also included. Different approaches to obtain N_e were gathered: $b(t)$ as described in (4), $b(\text{date})$ as described in Gutiérrez *et al.*, [11], $\log b(\text{date})$ as described in Pérez-Enciso [20], $\text{Var}(\text{offs})$ as described in (3), and $\overline{N_e}$ based on the individual increase in inbreeding together with its standard error. The theoretical N_e is also addressed in the table when possible. The method using the variances of family size does not result in accurate N_e estimates in examples reflecting scenarios closer to reality (such as population subdivision) because they only consider one source of variation of N_e . The N_e 's computed from regression approaches (regardless of whether it is done on generations or birth date) tend to perform poorly when reference subsets are defined in a pedigree file, particularly in extreme cases in which the subset of interest has lower mean F than the other individuals included in the pedigree file. Interestingly enough $\overline{N_e}$ is quite stable and precise regardless of the pedigree file (actual or simulated) analysed and the particular conditions of the whole and the reference subsets defined.

Table III. Effective population size computed in four simulated examples and a real population for the whole pedigree file and the last 400 individuals born (simulated examples) or the Carthusian reference subsets (RS, see text).

Pedigrees	Theoretical N		Whole pedigree file				Reference Subsets				
			b(t)	b(date)	logb(date)	Var(off)	b(t)	b(date)	logb(date)	$\overline{N_e}$	s.d.($\overline{N_e}$)
Simulated											
i) Random- two sexes	200.5	400	214.0	202.3	201.7	216.1	163.7	162.9	163.0	204.9	1.8466
ii) Subdivision	200.5/50.5	400	94.6	86.1	83.1	196.6	89.9	89.8	93.9	83.3	0.5917
iii) Overlapping	125	400	145.6	126.0	125.1	130.7	129.7	142.7	148.7	126.0	0.7420
iv) Non-Poisson	400	400	412.6	400.8	400.2	400.0	-11,671.6	-11,308.7	-3,749.1	401.6	6.4152
Carthusian											
Whole pedigree	-	6318	-	-	-	77.2	19.5	19.3	19.4	26.4	0.2563
RS ₇₉₋₈₉	-	1464	-	-	-	-	12.2	11.7	11.2	21.4	0.2067
RS ₁₀	-	1721	-	-	-	-	33.4	-49.8	-46.4	22.2	0.1817

Effective size obtained b(t) – via regression of inbreeding over equivalent generations , b(date) – via regression on birth date, logb(date) - via log regression on birth date, Var(off) – via variance of family size, and $\overline{N_e}$ - via individual increase in inbreeding. s.d. ($\overline{N_e}$) is the standard error of $\overline{N_e}$.

4. DISCUSSION

The simple methodology presented here to assess N_e in real populations accounts for pedigree knowledge of each individual in a population to obtain values of individual increase in inbreeding (ΔF_i). The increase in inbreeding is not treated here as a single value but as a variable with an associated mean ($\overline{\Delta F}$) that can easily be used to compute N_e (in fact $\overline{N_e}$) for a given reference subset

as $\overline{N_e} = \frac{1}{2\overline{\Delta F}}$. The current methodology addresses N_e directly from ΔF that,

theoretically, becomes constant in a population with stable size and breeding policy. This is contrary to F that increases from one generation to another [7]. Thus, ΔF is completely independent of pedigree depth in the idealized population [7]. The use of ΔF s additionally contributes to overcome the problem of using the F_i coefficients because the latter are non-linearly dependent on the pedigree depth of each individual. When trying to assess ΔF , after averaging F_i coefficients by generation, differences in absolute mean values from one generation to another must still be divided by one minus the mean inbreeding in the previous generation. This is not easy to carry out in real populations usually with overlapping. In such a scenario it is unrealistic to work under assumptions such as no inbreeding in previous generations, or linear trend of inbreeding by generations. However, individual increase in inbreeding is “free” from these effects since it is also adjusted for the generation number of each animal. Of course, ΔF s are still dependent on the completeness of the analysed pedigree file and need a few generations to become constant at the population level (Fig. 3). Notice also that in real-diploid populations in which self-fertilization is not possible and where there are two

different sexes, ΔF (but also $\overline{\Delta F}$), will need more generations to reach asymptotic values than those expected in the idealized population. In this respect the simple plot of ΔF_i over equivalent generations helps with the correct definition of the reference subsets and on its usefulness to obtain reliable estimates of N_e (Fig. 4).

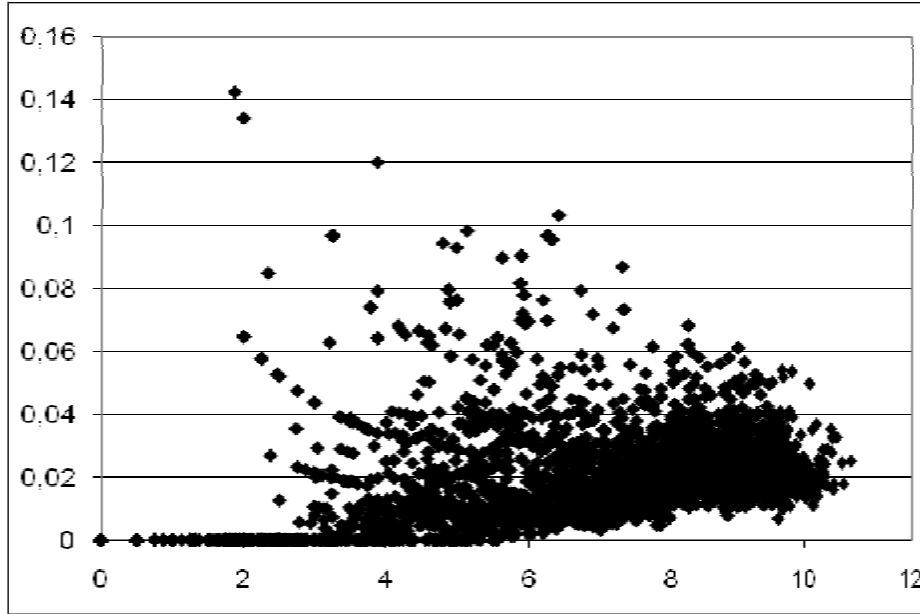


Figure 4. Plot summarising the dispersal of the individual increase in inbreeding (ΔF_i on the Y-axis) per individual number of equivalent generations (on the X-axis) in the whole Carthusian pedigree file.

Of course, as with other methodologies, the $\overline{N_e}$ obtained is still dependent on the way the reference subset is defined. However, the present approach can be used to assess the confidence interval of $\overline{N_e}$ from the standard error of $\overline{\Delta F}$. Indeed, no standard error can be computed directly for $\overline{N_e}$ because the presence in our reference subset of individuals with $\Delta F_i = 0$ would lead to individual estimates of $N_e = \infty$. Following the methodology presented here, values of $\Delta F_i = 0$ would affect $\overline{N_e}$ but an estimation of the parameter should still be possible.

When compared with the other methods to obtain N_e assayed here the N_e obtained from $\overline{\Delta F}$ is more stable regardless of the particular situation of the analysed pedigree files. This has been shown to be true regardless of whether we include in the reference subsets individuals with shallower or deeper pedigree (Tab. II). In this respect the use of the regression coefficient of the individual inbreeding values over the individual equivalent complete generations to compute N_e , gives estimates that are highly dependent on the mating policy carried out in the reference subset.

This last concern is particularly noticeable when, after having mated between relatives, a plan of matings based on low coancestry is implemented, leading to a decrease in the mean inbreeding thus providing negative N_e values.

As expected, the N_e computed from variances of family sizes is not useful to characterise the 'real' effective size, as shown in the Carthusian population analysed here. This is also accurately reflected by the results obtained from the simulated example ii) involving subdivision. If population subdivision is known, one would not treat a pedigree file as a single population, but each subpopulation separately. Then, the number of breeding animals would be smaller than when ignoring subdivision. The simulated example is only an unrealistic example of extreme subdivision but almost all real populations have some degree of uncontrolled subdivision, reflected by the fact that inbreeding is usually higher than coancestry. However, the method based on variance of family size reflects a temporary mating policy and can be useful when the pedigree knowledge is limited and/or subdivision has not occurred yet.

The methodology shown here does not consider phenomena such as population subdivision and selection. However, since these phenomena are reflected in the pedigree of individuals belonging to the analysed RS's, the effective size computed from ΔF_s is expected to capture these effects. The concept of effective size usually has an asymptotic meaning in a regular system and can be used for predictive purposes rather than for analysing genealogies [11]. The effective size obtained here is also related to pedigree tools used for describing genetic diversity in real populations, such as (effective) number of founders or ancestors [3, 15]. The parameters computed here ($\overline{\Delta F}$ and $\overline{N_e}$) are therefore related to those computed using the concept of long-term genetic contributions of ancestors [25, 26] even though the current approach does not need complex nor iterative computations. The relationship between inbreeding - actually the Lacy *et al.* [16] partial inbreeding coefficients - and founder contributions was used by Man *et al.* [19] to predict the frequency of carriers of an autosomal-deleterious gene when the ancestral source(s) of the gene is known. There is an inverse relationship between partial inbreeding coefficients coming from founders or ancestors and the effective number of founders or ancestors. In fact, under a random mating scheme, these partial inbreeding coefficients should be the sum of their squared contribution which is precisely the denominator in the computation of the effective number of founder or ancestors. The relationship between F_s and ΔF_s is straightforward and the inverse relationship between increase in inbreeding and N_e is given by definition.

However, the unbalance of the contributions from founders or ancestors is not the sole cause of the increases in inbreeding. They are also dependent on many other circumstances such as population structure, mating policy, changes in population size, etc. The estimates of N_e based on individual increase in inbreeding would accurately reflect the genetic history of the populations, namely the size of their founder population, their mating policy or bottlenecks due to abusive use of

reproductive individuals. All these phenomena influence the pedigree of the individual and are therefore reflected in the individual increase in inbreeding.

Inbreeding coefficients is widely used to calculate the rate of inbreeding and consequently N_e [7]. However, N_e can also be computed from the average coancestry of a RS [4]. In regular non-structured populations, average coancestry and inbreeding coefficients are analogous and the N_e obtained using either one approach should be the same. However, it has been reported that this is not always the case [23] and our results from the simulated population ii) would confirm this fact. The effective size is defined here from the individual rate of inbreeding. However, the method can be easily extended to apply the same methodology to coancestry coefficients. To do this, all the coancestries among individuals in the reference subset would be assigned to the inbreeding coefficient of a hypothetical offspring from each couple of individuals with an equivalent discrete generation number of half the sum of the parents plus one, by simply applying (3).

To conclude, we showed here a simple new approach to estimate N_e in real populations that give stable estimates for N_e . This methodology is derived from the definition of the individual increase in inbreeding. It is treated as a variable which can be computed for each individual and can be useful for other purposes. Moreover, the possibility of estimating N_e using this approach, and after assuming that it is the true N_e value, provides the chance to estimate other useful parameters by just making use of the multiple existent expressions to predict N_e in different situations [4, 7], as for examples rates of selection, migration and Hardy-Weinberg deviations. Although this parameter is still a novelty, its usefulness makes it worth while to investigate it further.

ACKNOWLEDGMENTS

The authors wish to thank the 'Subdirección General de Medios de Producción Ganaderos' of the Spanish Ministry of Agriculture for their support. We also thank the 'Fondo de Explotación de los Servicios de Cría Caballar y Remonta' of the Spanish Ministry of Defence and the Horse Breeders' Association for their collaboration. This work was partially funded by a grant from the Junta de Andalucía given to I. Cervantes (BOJA, nº. 120, 21/06/2004).

REFERENCES

- [1] Álvarez I., Royo L.J., Gutiérrez J.P., Fernández I., Arranz J.J., Goyache F., Relationship between genealogical and microsatellite information characterising losses of genetic variability: empirical evidence from the rare Xalda sheep breed, *Livest. Sci.* 115 (2008) 80-88
- [2] Baumung R., Sölkner J., Pedigree and marker information requirements to monitor genetic variability, *Genet. Sel. Evol.* 35 (2003) 383-369.
- [3] Boichard D., Maignel L., Verrier, E., The value of using probabilities of gene origin to measure genetic variability in a population, *Genet. Sel. Evol.* 29 (1997) 5-23.
- [4] Caballero A., Developments in the prediction of effective population size, *Heredity* 73 (1994) 657-679.

- [5] Caballero A., Toro M.A., Interrelations between effective population size and other pedigree tools for the management of conserved populations, *Genet. Res.* 75 (2000) 331-343.
- [6] Duchev Z., Distl O., Groeneveld E., Early warning system for loss of diversity in European livestock breeds, *Archiv. Anim. Breed.* 49 (2006) 521-531.
- [7] Falconer D.S., Mackay F.C., Introduction to quantitative genetics, 4th edn, Longman Group Ltd., England, 1996.
- [8] FAO, Secondary guidelines for the national farm animal genetic resources management plans: management of small populations at risk, FAO, Rome, Italy, 1998.
- [9] González-Recio O., López de Maturana E., Gutiérrez J.P., Inbreeding depression on female fertility and calving ease in Spanish dairy cattle, *J. Dairy Sci.* 90 (2007) 5744–5752.
- [10] Goyache F., Gutiérrez J.P., Fernández I., Gómez E., Álvarez I., Díez J., Royo L.J., Monitoring pedigree information to conserve the genetic variability in endangered populations: the Xalda sheep breed of Asturias as an example, *J. Anim. Breed. Genet.* 120 (2003) 95-103.
- [11] Gutiérrez J.P., Altarriba J., Díaz C., Quintanilla R., Cañón J., Piedrafita J., Pedigree analysis of eight Spanish beef cattle breeds, *Genet. Sel. Evol.* 35 (2003) 43-64.
- [12] Gutiérrez J.P., Goyache F., A note on ENDOG: a computer program for analysing pedigree information, *J. Anim. Breed. Genet.* 122 (2005) 357-360.
- [13] Hill W.G., A note on effective population size with overlapping generations, *Genetics* 92 (1979) 317-322.
- [14] James J.W., A note on selection differentials and generation length when generations overlap, *Anim. Prod.* 24 (1977) 109-112.
- [15] Lacy R.C., Analysis of Founder Representation in Pedigrees: Founder Equivalent and Founder Genome Equivalents, *Zoo. Biol.* 8 (1989) 111-123.
- [16] Lacy R.C., Alaks G., Walsh A., Hierarchical analysis of inbreeding depression in *Peromyscus polionotus*, *Evolution* 50 (1996) 2187–2200.
- [17] Kempen, G.M.P., van Vliet, L.J., Mean and variance of ratio estimators used in fluorescence ratio imaging, *Cytometry* 39 (2000) 300-305.
- [18] Maignel L., Boichard D., Verrier E., Genetic variability of French dairy breeds estimated from pedigree information, *Interbull Bulletin* 14 (1996) 49-54.
- [19] Man W.Y.N., Nicholas F.W., James J.W., A pedigree-analysis approach to the descriptive epidemiology of autosomal-recessive disorders, *Prevent. Vet. Med.* 78 (2007) 262-273.
- [20] Pérez-Enciso, M., Use of the uncertain relationship matrix to compute effective population size, *J. Anim. Breed. Genet.* 112 (1995) 327-332.
- [21] Royo L.J., Álvarez I., Gutiérrez J.P., Fernández I., Goyache F., Genetic variability in the endangered Asturcón pony assessed using genealogical and molecular information, *Livest. Sci.* 107 (2007) 162-169
- [22] Valera M., Molina A., Gutiérrez J.P., Gómez J., Goyache F., Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain, *Livest. Prod. Sci.* 95 (2005) 57-66.

- [23] Wang, J., Effective Size and F-Statistics of Subdivided Populations. II. Dioecious Species, *Genetics* 146 (1997) 1465-1474.
- [24] Woolliams, J.A., Mäntysaari, E.A., Genetic contributions of Finnish Ayrshire bulls over four generations, *Anim. Sci.* 61 (1995), 177–187.
- [25] Woolliams J.A., Wray N.R., Thompson R., Prediction of long-term contributions and inbreeding in populations undergoing mass selection, *Genet. Res.* 62 (1993) 231-242.
- [26] Wray N.R., Thompson R., Prediction of rates of inbreeding in selected populations, *Genet. Res.* 55 (1990) 41-54.
- [27] Wright S., *Evolution and the Genetics of Populations: The Theory of Gene Frequencies*, Vol. 2 .University of Chicago Press, Chicago, USA, 1969.

4.4. ARTICULO 4: APPLICATION OF INDIVIDUAL INCREASE IN INBREEDING TO ESTIMATE REALISED EFFECTIVE SIZES FROM REAL PEDIGREES

Journal of Animal Breeding and Genetics (2008), 125: 301-310

I. Cervantes^{1,3}, F. Goyache², A. Molina³, M. Valera⁴ and J. P. Gutiérrez^{1*}

¹Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro s/n, E-28040-Madrid, Spain.

²SERIDA-Somío, C/ Camino de los Claveles 604, E-33203 Gijón (Asturias), Spain.

³Departamento de Genética, Universidad de Córdoba, Ctra. Madrid-Cádiz, km 396^a, E-14071 Córdoba, Spain.

⁴Departamento de Ciencias Agro-Forestales, EUITA, Universidad de Sevilla, Ctra. Utrera km 1, E-41013 Sevilla, Spain.

Received: 7 February 2008; Accepted: 30 May 2008

* Correspondence: J. P. Gutiérrez, Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro s/n, E-28040-Madrid, Spain. Tel and Fax: +34913943767. E-mail: gutgar@vet.ucm.es

Summary

The objective of this work was to test the performance of a recently proposed methodology for the estimation of realised effective size (N_e) based on individual increase in inbreeding (ΔF_i) on several real pedigrees: a) an experimental mice population; b) a closed pedigree of fighting bulls; c) the Spanish Purebred (SPB, Andalusian) horse pedigree; d) the Carthusian strain of SPB pedigree; e) the Spanish Arab horse pedigree; and f) the Spanish Anglo-Arab horse pedigree.

Several reference subpopulations were defined with the basis of generation length in order to consider only animals in the last generation, to assess the influence of the pedigree content on the estimates of N_e . The estimates of realised N_e computed from ΔF_i ($\overline{N_e}$) tended to be higher than those obtained from regression on equivalent generations. The new parameter $\overline{N_e}$ remained approximately stable when pedigree depth achieved about five equivalent generations. Estimates of $\overline{N_e}$ take into account the genetic history of the populations, the size of their founder population, and the mating policy or bottlenecks due to poorly planned use of reproducing individuals. The usefulness of the realised N_e computed from individual increase in inbreeding in real pedigrees is also discussed.

Keywords: Effective size; increase in inbreeding; overlapped generation

Introduction

The size of an idealized population, which would give rise to the rate of inbreeding, or the rate of change in variance of gene frequencies observed in the population under consideration defines the effective population size (N_e) (Wright, 1969). This concept is a key parameter widely used as a criterion for characterising the risk status of livestock breeds (FAO, 1998; Duchev *et al.*, 2006). This is basically due to the direct relationship between N_e and the rate of inbreeding, fitness and the amount of genetic variation lost due to random genetic drift (Falconer and Mackay, 1996; Caballero and Toro, 2000) which makes it useful to infer the future of genetic stock.

However, the theoretical assumptions on which N_e is based only fit with idealized populations (Caballero, 1994) which is extremely important in the development of methods enabling the computation of effective size from pedigrees of actual populations under generation overlapping, selection and no-random mating. A quick overview of recent literature shows that the generally accepted methods to compute N_e in real pedigrees are based on regressing the inbreeding coefficients on birth dates or generations (Pérez-Enciso, 1995; Gutiérrez *et al.*, 2003) even though they are always dependent on the definition of a reference population within pedigree and further adjustment by the generations interval (L). Moreover, they have not been proven to ascertain a reliable value when there are variations in the breeding policy affecting mean inbreeding in the reference population (Gutiérrez *et al.*, 2008). In such a scenario planning matings to minimize coancestry after a period in which mating between close relatives was preferred, the regression-based methods to compute effective size can give unrealistic negative N_e values (Gutiérrez *et al.*, 2008).

Many attempts have been made to deal with the different real world scenarios in order to obtain reliable estimates of the effective population size (Caballero, 1994; Caballero and Toro, 2000). However, there is not a standard method for general application to obtain the effective population size.

Gutiérrez *et al.* (2008) have recently described the properties of a method for the estimation of effective size from pedigrees based on the definition of the individual increase in inbreeding coefficient (ΔF_i ; González-Recio *et al.*, 2007) using simulated standard pedigrees and in a unique real pedigree.

The objective of this work was to test the methodology for estimating N_e based on ΔF_i in several real pedigrees of varying depth in pedigree depth, overlapping generations, subdivision, migration, unbalanced representation of founders, or other unknown effects affecting the breeding policy, in order to further establish the usefulness of this parameter in real livestock breeding scenarios. To achieve these goals data from an experimental mice population, the Domecq fighting bull herd pedigree, the Spanish Purebred horse (SPB) pedigree, the Carthusian strain of SPB pedigree, the Spanish Arab horse pedigree and the Spanish Anglo-Arab horse pedigree were used to test the methodology developed by Gutiérrez *et al.* (2008). A comparison of the obtained estimates with other well-known methods for computing N_e will be done.

Materials and methods

Individual increase in inbreeding

The original idea of the methodology is to assume for each individual that all the animals in the referred population are under the same situation regarding inbreeding and genealogical information; this is measured over t , the discrete equivalent generation (Maignel *et al.*, 1996, Boichard *et al.*, 1997). The discrete generation equivalent is computed for the pedigree of each individual as the sum over all known ancestors of the term of $(\frac{1}{2})^n$, where n is the number of generations separating the individual from each known ancestor. The steady increase in inbreeding that would have led to the inbreeding coefficient of animal i (F_i) in t generations should be:

$$\Delta F_i = 1 - \sqrt[t]{1 - F_i} \quad (1)$$

Following Gutiérrez *et al.* (2008), an overall mean of this parameter ($\overline{\Delta F}$) can be easily computed by averaging the ΔF_i s of the individuals included in a defined reference subpopulation. From this, as a 'realised' effective population size, $\overline{N_e}$

can be straightforwardly computed as $\overline{N_e} = \frac{1}{2\overline{\Delta F}}$. The realised effective size

should be understood as the global effective size over time which has led to the present inbreeding level in the reference population from the founder population.

Moreover, given that an individual increase in inbreeding exists for each animal in the reference subpopulation, a standard error of the realised $\overline{N_e}$ can be further computed from the standard deviation of the individual increase in inbreeding and the square root of the reference subpopulation size as:

$$\sigma_{\overline{N_e}} = 2\overline{N_e}^2 \sigma_{\Delta F} \frac{1}{\sqrt{\overline{N_e}}}.$$

This is a modification of the expression reported by Gutiérrez *et al.* (2008):

$$\sigma_{\overline{N_e}} = 2\overline{N_e}^2 \sigma_{\Delta F} \frac{1}{\sqrt{n}}$$

where n has been changed to $\overline{N_e}$, given that this value would be an equivalent number of independent individuals instead of n . Note also that this standard error gives an approximate idea of certainty but has a limited informative value given that the distribution of ΔF_i is skewed to the right.

Other methods to estimate N_e using pedigree information

Several approaches have been used to check estimates of N_e obtained from individual increase in inbreeding:

1) First, N_e was estimated from the variances of family size as

$$\begin{aligned} \frac{1}{N_e} = \frac{1}{16ML} & \left[2 + \sigma_{mm}^2 + 2\left(\frac{M}{F}\right)\text{cov}(mm, mf) + \left(\frac{M}{F}\right)^2 \sigma_{mf}^2 \right] \\ & + \frac{1}{16FL} \left[2 + \left(\frac{F}{M}\right)^2 \sigma_{fm}^2 + 2\left(\frac{F}{M}\right)\text{cov}(fm, ff) + \sigma_{ff}^2 \right] \end{aligned} \quad (\text{Hill, 1979})$$

where M and F are the number of male and female individuals born or sampled for breeding each time period, L the average generation interval, σ_{mm}^2 and σ_{mf}^2 are the variances of the male and female offspring of a male, σ_{fm}^2 and σ_{ff}^2 are the variances of the male and female offspring of a female, $\text{cov}(mm, mf)$ and $\text{cov}(fm, ff)$ the respective covariances. Note that the family size of a parent (male or female) consists of sons and daughters kept for reproduction (James, 1977). This should be a good method when all the other conditions are in compliance with the idealized population but generations are overlapping (Falconer and Mackay, 1996).

2a) Moreover, N_e was also estimated from ΔF following Gutiérrez *et al.* (2003). The increase in inbreeding between two generations ($F_t - F_{t-1}$) was obtained from the regression coefficient (b) of the inbreeding coefficient over the date of birth obtained in the reference subpopulation, and considering the average generation

interval (L) as $F_t - F_{t-1} = L \times b$ with F_{t-1} computed from the mean inbreeding in the reference subpopulation (F_t) as: $F_{t-1} = F_t - L \times b$.

2b) N_e was also computed in a very similar way as 2a but using directly t instead of approaching the generations through generation intervals. Using this approach, N_e was computed from the regression coefficient (b) of the individual inbreeding values over the individual equivalent complete generations t . In this

case, $\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}} \approx \frac{b}{1 - (F_t - b)}$, where F_t the average F of the reference subpopulation.

3) Estimation of N_e could also be approached by using

$1 - F_t = \left(1 - \frac{1}{2N_e}\right)^t$ to derive its value from a log regression of $(1 - F)$ over generation number (Pérez-Enciso *et al.*, 1995). Here we computed the log regression of $(1 - F)$ over birth date and corrected further by the length of the generation interval (Pérez-Enciso *et al.*, 1995).

Examples

To assess this methodology it was applied to several pedigrees in common scenarios in real populations:

- a) An experimental mice population (Ibáñez-Escriche *et al.*, 2008; Moreno, 1998) consisting of 18 discrete generations with about 72 males and 144 females per generation structured in 9 replicates of selection. Mating policy was different in three replicates of each of the selection method: at random, weighted selection, and minimum coancestry.
- b) The pedigree information available from the beginning of the 20th century in the Fighting Bull herd of Juan Pedro Domecq, including a total of 33,330 individuals.
- c) The pedigree of Spanish Purebred horse (SPB, Andalusian horse) from their foundation in the XIXth century until December 2004 (Ministry of Defense: www.librogenealogico.com). The total available records were 140,629.
- d) The pedigree of the Carthusian strain of the Spanish Purebred horse, which is a subset of the pedigree listed before having a particular management, including those founder individuals classified as Carthusian due to their particular historical origin and those individuals resulting from matings between Carthusian parents (Valera *et al.*, 2005). This is the example population already used for this purpose by Gutiérrez *et al.* (2008), but has

also been included here as a quick comparison among the different methodologies and pedigrees.

- e) The pedigree of the Spanish Arab horse recently analysed by Cervantes *et al.* (2008), including a total of 18,880 individuals.
- f) The pedigree of the Spanish Anglo-Arab horse, including a total of 10,568 individuals. The data file was obtained from the Ministry of Defence (web site www.librogenealogico.com).

The main parameters describing the real pedigrees analysed are given in Table I. The examples listed above are expected to characterise a classical example of population with discrete-inbred generations (mice), closed population with overlapping generations and matings between relatives (SPB, Carthusian strain), closed population with overlapping generations and matings planned to avoid inbreeding (Fighting Bull) and, finally, open populations with different pedigree depths varying from 'sufficiently' deep (Spanish Arab horse) to very shallow (Spanish Anglo-Arab horse) as a consequence of different rates of migration. The mice dataset also included a strong population subdivision in 9 independent lines from the foundation of the population (Ibáñez-Escriche *et al.*, 2008; Moreno, 1998). The Carthusian population was reported to be managed using minimum coancestry mating policies during recent generations (Valera *et al.*, 2005) and the Fighting Bull population is a case of a population of constant size with a mating policy avoiding common ancestors in the last three generations of each individual.

Several reference subpopulations were defined trying to take into account the last generation interval, and also trying to assess the influence of the pedigree content on the estimates of N_e . For the mice pedigree the 15-16th and the 18th generations were used as two reference subpopulations (RS). For the other pedigrees that were analyzed various RS were defined according to three different strategies: a) pedigree depth of the individuals; b) the individuals born in the last 10 years of the available records; and c) individuals born in a given period of years allowing their use for reproduction. RS included individuals with 4 equivalent generations or more, and 8 equivalent generations or more except for the Spanish Anglo-Arab horse for which only the first RS was defined, given their extremely shallow pedigree.

Table 1. Number of individuals (N), average number of equivalent generations and standard deviation ($t \pm$ s.d.), maximum number of equivalent generations (Max. t), average inbreeding (F_i in percentage), average individual inbreeding rate (ΔF_i in percentage) and number of male and female reproductive individuals and average family size for males (A_m) and females (A_f) for the whole (WP) and the reference (RS) populations used as examples in the present analyses.

Population	N	$t \pm$ s.d.	Max. t	F (%)	ΔF (%)	Males	A_m	Females	A_f
Mice									
WP	9,152	9.2 \pm 5.25	18.3	12.6	1.27%	1,395 ^a	2.9	2,243 ^b	1.8
RS ₁₈	283	17.9 \pm 1.46	18.3	26.2	1.67%				
RS ₁₅₋₁₆	985	15.7 \pm 0.71	16.3	24.0	1.74%	160 ^a	2.7	239 ^b	1.8
Fighting Bull									
WP	33,330	5.0 \pm 3.20	10.9	1.3	1.29%	348 ^a	15.3	3,072 ^b	1.9
RS ₁₀	5,663	9.0 \pm 0.91	10.9	1.8	1.82%	24 ^a	8.6	100 ^b	1.3
RS ₉₂₋₀₁	6,641	8.3 \pm 0.81	10.3	1.8	1.77%	51 ^a	11.3	392 ^b	1.5
SPB									
WP	140,629	9.0 \pm 1.82	11.8	8.2	0.95%	5,475 ^a	7.5	14,689 ^b	2.8
RS ₁₀	89,324	9.9 \pm 0.58	11.8	8.4	0.90%	210 ^a	2.2	738 ^b	1.1
RS ₈₃₋₉₃	55,737	9.3 \pm 0.67	11.3	8.8	1.00%	1,755 ^a	4.4	5,631 ^b	1.7
Carthusian									
WP	6,318	6.6 \pm 2.75	10.9	13.0	1.89%	424 ^a	4.9	933 ^b	2.3
RS ₁₀	1,721	9.1 \pm 0.68	10.9	18.6	2.25%	1 ^a	1.0	5 ^b	1.2
RS ₇₉₋₈₉	1,464	8.2 \pm 0.64	9.8	17.5	2.34%	97 ^a	4.1	167 ^b	2.1
Arab									
WP	18,880	5.7 \pm 2.78	10.3	7.0	1.07%	1,560 ^a	4.2	3,260 ^b	2.0
RS ₁₀	6,240	7.9 \pm 1.28	10.3	9.8	1.28%	22 ^a	1.8	36 ^b	1.1
RS ₈₆₋₉₇	5,557	7.1 \pm 1.60	9.7	9.7	1.37%	307 ^a	2.7	663 ^b	1.5
Anglo-Arab									
WP	10,568	1.7 \pm 1.51	6.2	0.8	0.24%	757 ^a	2.8	1,325 ^b	1.5
RS ₁₀	2,862	2.4 \pm 1.59	6.2	0.9	0.24%	2 ^a	1.0	5 ^b	1.2
RS ₈₄₋₉₆	2,920	1.7 \pm 1.64	5.6	0.9	0.28%	129 ^a	1.8	225 ^b	1.3

RS₁₈: animals in the 18th generation; RS₁₅₋₁₆: animals in the 15th and 16th generation; RS₁₀: animals born in the last decade. Other RS_{xx-yy}: animals born between the years 19xx and 19yy.

^aindividuals born in the defined period that acted sequentially as male parents.

^bindividuals born in the defined period that acted sequentially as female parents.

Program used

The analyses were performed using the ENDOG program (current version v4.5) (Gutiérrez and Goyache, 2005), which can be freely downloaded from the World Wide Web at http://www.ucm.es/info/prodanim/html/JP_Web.htm.

Results

Table 1 gives the main parameters describing the pedigrees that were analysed. Whole pedigreed set sizes varied from 6,318 (Carthusian) to 140,629 (SPB) whilst the sizes for the fitted RS varied from 283 individuals for RS_{18} in mice to 89,324 for RS_{10} in SPB. The mean equivalent generations ranged between 1.7 ± 1.51 for WP in Spanish Anglo-Arab horse to 17.9 ± 1.46 for RS_{18} in mice. The maximum value for equivalent generations were observed in the mice population (from 16.3 to 18.3), it was over 10-11 for the different RS fitted for Domecq's Fighting Bull, SPB, Carthusian and Spanish Arab horses and reached lower values (from 5.6 to 6.2) for the Spanish Anglo-Arab horse. The highest average inbreeding values at the WP level were found in Carthusians (13.0%) and the lowest (0.8%) in the population with the shallowest pedigree (Spanish Anglo-Arab horse). The lowest increase in inbreeding were found in the Spanish Anglo-Arab horse (0.2%) and in the SPB (lower than 1%) whilst the highest was found in the Carthusian subpopulation (around 2%).

Figure 1 shows the plot of the values of individual increase in inbreeding per number of equivalent generations in all the pedigrees analysed. Notice that the trend of the values of ΔF_i tended to stabilize with the increase in t (about five) due to the correction for pedigree depth of the individuals.

Figure 2 shows the evolution of the mean inbreeding, mean individual increase in inbreeding, mean equivalent generations, and realised effective size ($\overline{N_e}$) across years of birth (generations in the mice population) for all the pedigrees. It is worth pointing out that the mean individual increase in inbreeding became approximately stable in the last period of time, whilst mean inbreeding increased and mean equivalent generations linearly increased during this last period. Furthermore, $\overline{N_e}$ remains approximately stable when known pedigree reaches about five equivalent generations. This effect is not so clear in the Spanish Anglo-Arab population with lower pedigree knowledge. A decrease in mean individual increase in inbreeding (and a related increase in $\overline{N_e}$) was observed in the SPB population.

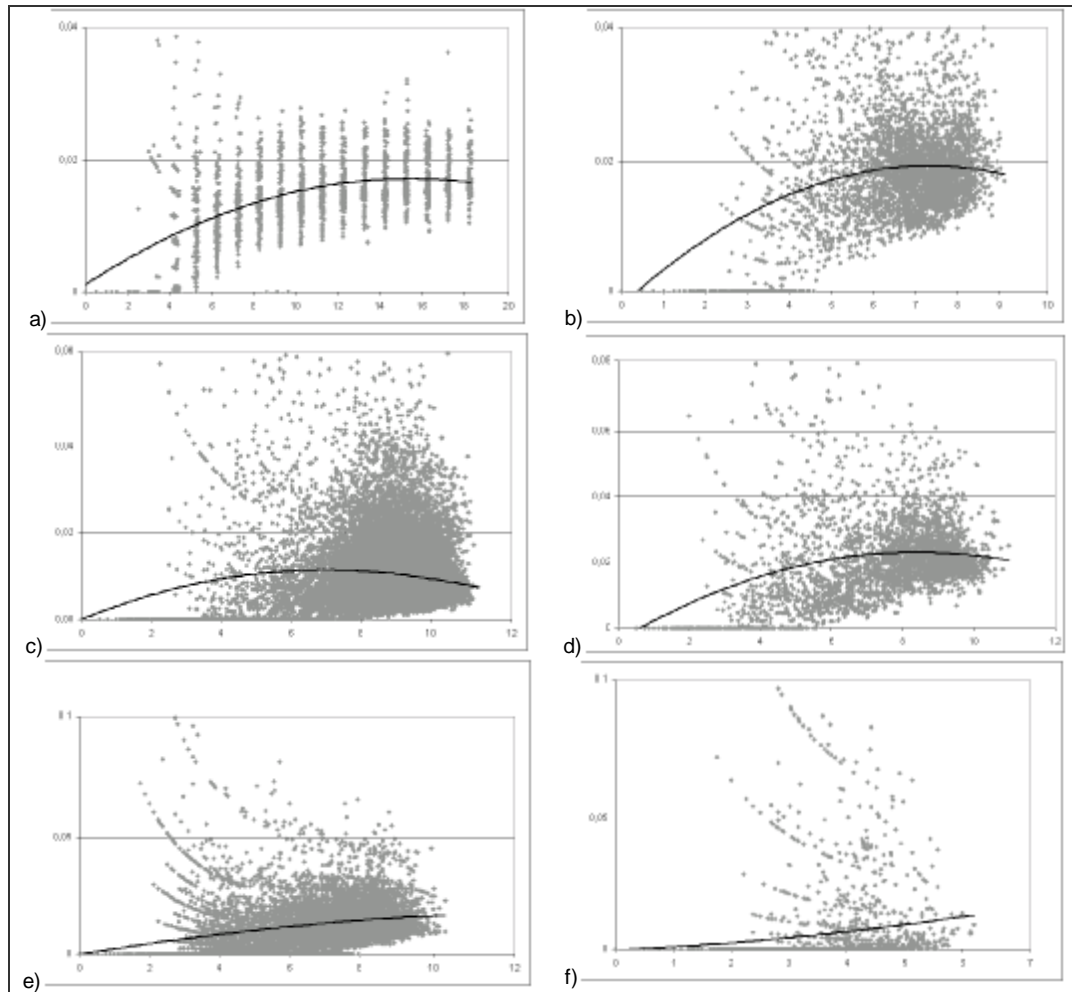


Figure 1. Plots summarising the dispersal and the trend line of the individual increase in inbreeding (ΔF_i on the Y-axis) per individual number of equivalent generations (on the X-axis) in the whole analysed pedigrees: a) mice population; b) Fighting Bull population; c) Spanish Purebred horse; d) Carthusian strain of the Spanish Purebred horse; e) Spanish Arab horse; f) Spanish Anglo-Arab horse.

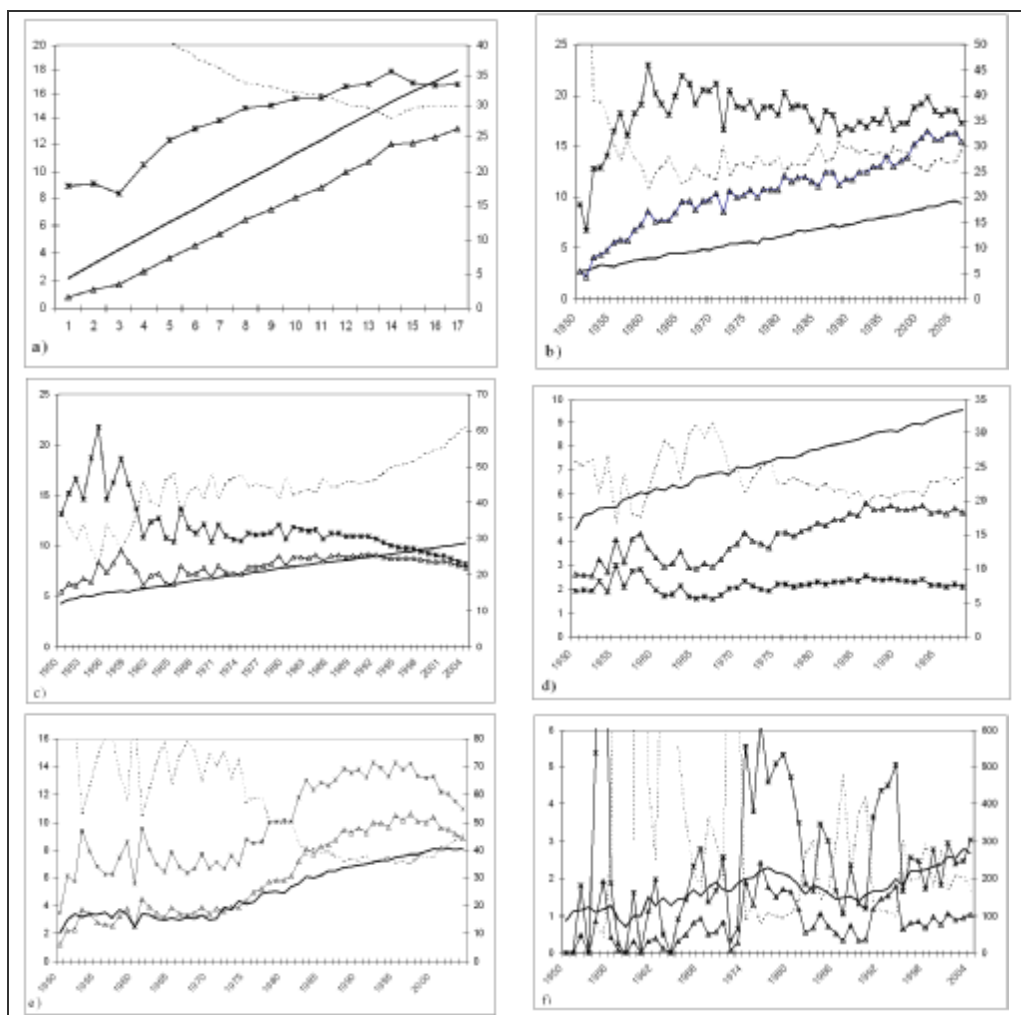


Figure 2. Evolution of the mean inbreeding (Δ , right axis in mice and Carthusian examples and left axis for the others) , mean individual increase in inbreeding ($\times 10$, except carthusian, $*$, left axis), mean equivalent generations ($-$, left axis), and N_e ($---$, right axis) across year of birth (across generations in mice), in the analysed pedigrees: a) mice population; b) Fighting Bull population; c) Spanish Purebred horse; d) Carthusian strain of the Spanish Purebred horse; e) Spanish Arab horse; f) Spanish Anglo-Arab horse.

Table 2 includes the estimates of N_e obtained using regression of the individual coefficients of inbreeding on equivalent generations, variance of family size and individual increase in inbreeding (ΔF_i) in the different RS analysed here. Also included in this table are restrictions in pedigree depth of the individuals included in the corresponding RS. Regardless of the RS being considered, the estimates obtained using variances in family size are larger than those obtained using both regression on equivalent generations and using ΔF_i except for the population with shallower pedigree (Spanish Anglo-Arab horse). In this latter population the regression coefficients are affected by the existence of a few highly inbred individuals whilst most individuals in the populations have $F = 0$ since about 30% of Spanish Anglo-Arab horse are founders. In this respect, the larger estimates for the obtained $\overline{N_e}$ using ΔF_i reflect more appropriately the genetic situation of such populations with shallow pedigrees than those based on regression of individual values of inbreeding on equivalent generations. A possible explanation for this is that the regression of individual values assumes a quadratic loss function (applying the largest weight on extreme values), while taking the mean of individual values assumes a linear loss function (applying equal weight on all values).

The estimates of realised N_e computed from $\Delta F_i (\overline{N_e})$ tend to be higher than those obtained from regression on equivalent generations. Moreover, when RS is analysed, the values of $\overline{N_e}$ are quite stable regardless of any restriction of the pedigree depth of the individuals included in the corresponding RS, whilst the N_e obtained from regression of F on equivalent generations show an undesired behavior. In the Carthusian population it is possible to find a noticeable increase of the N_e value from 20.4 to 40.0 when RS₁₀ includes individuals with 4 or more equivalent generations in the pedigree but also negative values of N_e (-142.2) when RS includes individuals with 8 or more equivalent generations in the pedigree. This was similar to that assessed in the SPB, where N_e ranged between 147.4 and -533.1 for WP. Note that a negative estimate of N_e can be obtained when the increase in inbreeding is attained by regression of the inbreeding coefficient on the year of birth and younger individuals are less inbred than the older. Thus, the N_e obtained depends partially, on the effect of the changes in the mating policy. The criterion of minimal coancestry has recently been used to define the mating policy (Valera *et al*, 2005) in the Carthusian and SPB populations. However, $\overline{N_e}$ shows a more stable value of about 22 for the Carthusian population and 52 for the SPB population.

Table 2. Estimates of N_e using the whole pedigrees (WP) and reference subpopulations fitted using the individuals born in the last 10 years (RS₁₀) or in a given period of years (RS_{xx-yy}) obtained using regression of the individual coefficients of inbreeding on equivalent generations [b(t)], variance of family size (V) and individual increase in inbreeding ($\overline{N_e}$). Analyses were carried out without pedigree depth restrictions and using only individuals with 4 equivalent generations or more ($t \geq 4$) and 8 equivalent generations or more ($t \geq 8$).

Population	No pedigree depth restriction			$t \geq 4$			$t \geq 8$		
	N_e [b(t)]	N_e (V)	$\overline{N_e} \pm$ s.d.	N_e [b(t)]	$\overline{N_e} \pm$ s.d.	N	N_e [b(t)]	$\overline{N_e} \pm$ s.d.	N
Mice									
WP	27.1	198.4	39.4 \pm 5.66	24.8	34.2 \pm 2.42	7,307	24.2	31.3 \pm 1.52	5,289
RS ₁₈	12.3		29.9 \pm 1.33	12.3	29.9 \pm 1.33	283	12.3	29.9 \pm 1.33	283
RS ₁₅₋₁₆	19.7	223.1	28.8 \pm 1.21	19.7	28.8 \pm 1.21	985	19.7	28.8 \pm 1.21	985
Fighting Bull									
WP	25.0	22.2	38.8 \pm 6.16	29.0	26.9 \pm 2.23	20,482	26.1	27.2 \pm 1.30	7,654
RS ₁₀	19.0		27.5 \pm 1.46	17.8	27.5 \pm 1.43	5,648	33.4	26.9 \pm 1.20	5,300
RS ₉₂₋₀₁	17.7	101.1	28.2 \pm 1.70	16.5	28.1 \pm 1.67	6,624	24.1	27.3 \pm 1.38	4,776
SPB									
WP	64.0	257.9	52.4 \pm 6.01	147.4	51.2 \pm 5.31	135,813	-533.1	52.3 \pm 4.92	121,371
RS ₁₀	146.4		55.4 \pm 4.96	150.9	55.4 \pm 4.96	89,318	172.3	55.4 \pm 4.94	89,080
RS ₈₃₋₉₃	62.9	2570.6	50.1 \pm 4.79	67.9	50.0 \pm 4.78	55,681	85.0	50.0 \pm 4.72	54,722
Carthusian									
WP	19.5	32.5	26.4 \pm 3.96	20.4	22.4 \pm 2.64	5,172	273.5	21.4 \pm 1.54	2,518
RS ₁₀	33.4		22.2 \pm 1.60	40.0	22.2 \pm 1.59	1,719	-142.2	22.0 \pm 1.53	1,671
RS ₇₇₋₈₉	12.2	88.1	21.4 \pm 1.71	11.6	21.3 \pm 1.71	1,463	21.9	20.3 \pm 1.49	924
Arab									
WP	30.6	121.7	46.6 \pm 7.04	22.0	37.7 \pm 3.93	13,624	21.6	33.1 \pm 2.67	4,138
RS ₁₀	18.1		39.1 \pm 4.19	15.3	38.4 \pm 4.00	6,093	19.2	33.8 \pm 2.80	3,495
RS ₈₆₋₉₇	22.2	569.9	36.5 \pm 3.85	16.2	34.5 \pm 3.13	5,095	11.4	30.8 \pm 2.27	1,364
Anglo-Arab									
WP	71.1	142.7	210.0 \pm 72.30	870.9	61.6 \pm 13.81	1,022			
RS ₁₀	80.5		209.8 \pm 68.20	91.2	84.7 \pm 19.37	612			
RS ₈₄₋₉₆	74.1	372.8	175.8 \pm 59.70	25.5	96.8 \pm 20.34	255			

RS₁₈: animals in the 18th generation; RS₁₅₋₁₆: animals in the 15th and 16th generation; RS₁₀: animals born in the last decade. Other RS_{xx-yy}: animals born between the years 19xx and 19yy; N: number of individuals.

Discussion

The estimation of the realised N_e from individual increase in inbreeding has been shown to perform reasonably well in simulated examples (Gutiérrez *et al.*, 2008) from the comparison between expected and estimated values. Its performance in

real populations is rather complex since the real value of the parameter is unknown and no point of comparison exists. To assess this performance N_e has been computed from individual increase in inbreeding in several real populations analysing the influence of the definition of different RS relating to more or less recent periods and pedigree knowledge. Some properties of the realised N_e computed from individual increase in inbreeding were tested in these real pedigrees.

This methodology addresses the realised N_e directly from ΔF that theoretically becomes constant in an ideal population instead of addressing it from F that increase from one generation to another in the idealized population (Falconer and Mackay, 1996). By definition ΔF is thus completely independent of pedigree depth in the idealized population (Falconer and Mackay, 1996). Therefore, the use of ΔF_i contributes to overcome the problem of using the F_i coefficients because the latter are non-linearly dependent on the pedigree depth of each individual. Of course, ΔF_i are still dependent on the completeness of the analysed pedigree and need a few generations to become constant at the population level (Figures 1 and 2). In this respect, the simple plot of ΔF_i over generations serves as a guide for correctly defining RS to obtain a reliable estimate of the realised N_e (Figure 1). The definition of RS is still critical to obtain useful values of N_e , but this approach makes it possible to assess the confidence level of $\overline{N_e}$ from the standard deviation of $\overline{\Delta F}$ (Gutiérrez *et al.*, 2008).

Regression of inbreeding values based on time or generation to compute N_e can be useful when the aim is to know the influence of temporal mating policies in some reference subpopulations because this method provides estimates that are highly dependent on those policies. This last concern is particularly noticeable when, after mating relatives, a new mating strategy, based on low coancestry, is implemented, which leads to a decrease in the mean inbreeding, resulting in negative N_e values. On the contrary, when we compare $\overline{N_e}$ obtained from $\overline{\Delta F}$ with that obtained using other methods, $\overline{N_e}$ appears more stable regardless of the particular situation and depth of the analysed pedigrees (Table 2). In fact, $\overline{N_e}$ becomes a final measure of the accumulated drift, originating from the foundation of the population until the present, because it is unable to accommodate the consequences of a recent change in the mating policy, especially if the pedigree depth is high. If the aim for estimating the effective population size is to ascertain the effect of changes in the mating policy, it would be recommendable to truncate the pedigree, thus considering as founders those animals which were used to establish the new mating policy. However, always make sure that the depth of the remaining pedigree is enough to achieve reliable estimates.

The estimates of N_e computed from variances of family size are not useful to characterise the 'real' effective size of the population being studied (Gutiérrez *et al.*, 2008). This is clearly reflected by the results obtained from the mice population

analysed here. The values obtained using this method are approximately the real number of reproductive individuals in each mice generation (Table 2) because during the experiment the number of individuals that were selected to be parents of the following generation tended to be balanced across families. However, the value of N_e obtained using this method did not account for the structure of the population that has a permanent subdivision in 9 different replicates per generation. However, the analysed fighting bull population is managed under a minimum coancestry mating policy that tends to avoid subdivision. As a consequence, the value of N_e obtained using variances of family size is fairly reasonable for this population. Known subdivision could be taken into account, but this is an uncommon situation in non-idealized situations. Thus, this method only uses differences in family size to estimate N_e , ignoring the subdivision of the population which, in fact, causes the reduction of N_e regarding the number of reproductive animals and therefore, tend to overestimate the parameter. A second important reason explaining why this approach fails in predicting N_e is that it ignores the part of the selection process that causes correlations across generations. The Hill (1979) formula ignores the fact that animals with the highest number of selected sibs are also the animals most likely to have offspring selected.

Besides the number and unbalanced contributions of founders or ancestors, there are several causes that influence the effective size, for example population structure, mating policy, changes in population size, etc. This is shown by some of the databases analysed here. The effective number of ancestors computed for the Fighting Bull population (data not shown) is the lowest but its N_e is not so; this is a case of a population of constant size where there is a mating policy, which avoids common ancestors in the last three generations of each animal. A similar N_e is obtained for the mice population with much higher effective number of ancestors and moreover, a balanced contribution of parents to the next generation and, in some cases, with a minimum coancestry mating policy. However, the population is structured in 9 lines and N_e is practically identical to the real number of reproductive animals within lines. $\overline{N_e}$ explains the number of individuals that would give rise to the rate of inbreeding (Falconer and Mackay, 1996) when addressed from inbreeding coefficients, but when subdivision is present, the mean coancestry is lower than mean inbreeding and the Hill method should be preferred when a global idea of inbreeding is desired. Thus, the Hill method guesses better the real number of breeding individuals in the mice example. These results show that estimates of $\overline{N_e}$ take into account the genetic history of the populations. In particular the size of their founder population, their mating policy, bottlenecks due to abusive use of reproductive individuals, but is less affected by, for example, changes in the mating policy, or the accumulation of pedigree. On the other hand, realised effective population size has to be understood in the context of pedigree knowledge. If for instance, a bottleneck has occurred in the middle of a pedigree, it will affect the genetic stock only from this point in the past whilst the method also accounts from this point onwards. Then, if a breeder wishes to know the effective size in a population in a specific period, he should define the reference population

at the end of the desired period, and truncate the pedigree in the beginning of the period; always making sure that there are enough generations considered within the period to obtain a reliable estimate.

Comparison of $\overline{N_e}$ with other estimates of N_e is problematic since the real value is unknown. Looking at Figures 1 and 2 it can be noted that the individual increases in inbreeding becomes stable after some generations of pedigree depth. Thus, we have compared the $\overline{N_e}$ values with others based on regression or log regression over generations or dates of birth when only individuals with more than 4 equivalent generations known are considered (Table 3). Estimates of N_e based on regressions using the whole pedigreed set are also shown in Table 3.

Table 3. Estimates of N_e obtained using regression of the individual coefficients of inbreeding on equivalent generations [b(t)], regression on birth date, log regression on birth date and individual increase in inbreeding ($\overline{N_e}$).

	Whole Pedigree set				Reference Subpopulations (t ≥ 4)				
	N	b(t)	b (date)	Log b (date)	N	b(t)	b (date)	Log b (date)	$\overline{N_e} \pm \text{s.d.}(\overline{N_e})$
Mice	9,152	27.1	28.1	27.3	7,307	24.8	26.1	25.4	34.2 ± 2.42
Fighting Bull	33,330	25.0	28.1	27.7	20,482	29.0	30.6	30.6	26.9 ± 2.23
SPB	140,629	64.0	72.8	75.2	135,813	147.4	299.3	396.2	51.2 ± 5.31
Carthusian	6,318	19.5	19.3	19.4	5,172	20.3	23.0	23.6	22.4 ± 2.64
Arab	18,880	30.6	37.3	36.6	13,624	22.0	43.4	42.5	37.7 ± 3.93
Anglo-Arab	10,568	71.1	343.6	316.3	354	25.5	23.6	22.1	61.6 ± 13.81

Analyses were carried out using the whole pedigrees (WP) and reference subpopulations (RS), including only the individuals with 4 equivalent generations (t).

In general, the values of $\overline{N_e}$ obtained in the reference subpopulation are in close agreement with those of N_e obtained via regression (Table 3). In fact, values of $\overline{N_e}$ seem more reliable than the others when partial disagreement exists if a visual inspection of the evolution of N_e is used (Figure 2), but some of these values are worth commenting on. For instance, in the mice population, given that minimum coancestry is carried out for three of the replicates while random mating is in the other six replicates, stabilization of $\overline{N_e}$ is achieved in the latter generations and values lower than 30 are rare. $\overline{N_e}$ reaches a value of 34.2 while all the N_e values obtained by regression remain lower than 30 (Table 3). The case of SPB is also worth discussing. The change in the breeding policy brings a reduction in the average inbreeding of the last time points and generations with a consequent increase in N_e values. Looking at the evolution of N_e throughout the time points these values never exceeded more than 50 (Figure 2) and a $\overline{N_e}$ value of 51.2,

whereas the lowest value based on regression is 147.4 (Table 3). This seems to be because the regression is more sensitive to the decline in average inbreeding in recent years. A mention is also needed in the case of the two populations subject to migration in a low rate (Spanish Arab horse) or a high rate (Spanish Anglo-Arab horse), in both of them the increase in pedigree knowledge is not linear showing changes in the policy of migration. Again the evolution of N_e is unstable especially for the Spanish Anglo-Arab breed with very low pedigree knowledge. Thus, if considering N_e based on regression over time for the whole pedigreed set the values are higher than 300, but if only individuals with more than 4 equivalent generations are considered, the same methodology provides a value close to 20. Estimation of $\overline{N_e}$ results in 61.6, a coherent value but with a standard error much higher than those obtained for other populations.

To conclude, we tested here the new approach to estimate the realised N_e in real populations thus enabling us to obtain stable estimates of N_e .

Acknowledgments

The authors wish to thank the ‘Subdirección General de Medios de Producción Ganaderos’ of the Spanish Ministry of Agriculture for their support. The authors are indebted to the group ‘Modelos de Selección y Evaluación de Reproductores’ of the U.C.M. and to J.P. Domecq and to the staff of the Fighting Bull Breeder Association (UCTL) for providing access to their pedigree data. We also thank the ‘Fondo de Explotación de los Servicios de Cría Caballar y Remonta’ of the Spanish Ministry of Defence and the Horse Breeders’ Association for their collaboration. This work was partially funded by a grant from the Junta de Andalucía given to I. Cervantes (BOJA, nº. 120, 21/06/2004). The Authors also wish to express their sincere thanks and acknowledgement to one of the referees for their significant contributions to improve this paper.

References

- Boichard, D., Maignel, L., Verrier, E. (1997) The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Sel. Evol.*, **29**, 5-23.
- Caballero, A. (1994) Developments in the prediction of effective population size. *Heredity*, **73**, 657-679.
- Caballero, A., Toro, M.A. (2000) Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genet. Res.*, **75**, 331-343.
- Cervantes, I., Molina, A., Goyache, F., Gutiérrez, J.P., Valera, M. (2008) Population history and genetic variability in the Spanish Arab Horse assessed via pedigree analysis. *Livest. Sci.*, **113**, 24-33.
- Duchev, Z., Distl, O., Groeneveld, E. (2006) Early warning system for loss of diversity in European livestock breeds. *Archiv. Anim. Breed.*, **49**, 521-531.
- Falconer, D.S., Mackay, F.C. (1996) *Introduction to quantitative genetics*. 4th edn, Longman Group Ltd., England.

- FAO (1998) Secondary guidelines for the national farm animal genetic resources management plans: management of small populations at risk. FAO, Rome, Italy.
- González-Recio, O., López de Maturana, E., Gutiérrez, J.P. (2007) Inbreeding depression on female fertility and calving ease in Spanish dairy cattle. *J. Dairy Sci.*, **90**, 5744-5752.
- Gutiérrez, J.P., Altarriba, J., Díaz, C., Quintanilla, R., Cañón, J., Piedrafita, J. (2003) Pedigree analysis of eight Spanish beef cattle breeds. *Genet. Sel. Evol.*, **35**, 43-64.
- Gutiérrez, J.P., Cervantes, I., Molina, A., Valera, M., Goyache, F. (2008) Individual increase in inbreeding allows estimating effective sizes from pedigrees. *Genet. Sel. Evol.*, **40**, 359-378.
- Gutiérrez, J.P., Goyache, F. (2005) A note on ENDOG: a computer program for analysing pedigree information. *J. Anim. Breed. Genet.*, **122**, 357-360.
- Hill, W.G. (1979) A note on effective population size with overlapping generations. *Genetics*, **92**, 317-322.
- Ibáñez-Escriche, N., Moreno, A., Nieto, B., Piqueras, P., Salgado, C., Gutiérrez, J.P. (2008) Genetic parameters related to environmental variability of weight traits in a selection experiment for weight gain in mice; signs of correlated canalised response. *Genet. Sel. Evol.*, **40**, 279-293.
- James, J.W. (1977) A note on selection differentials and generation length when generations overlap. *Anim. Prod.*, **24**, 109-112.
- Maignel, L., Boichard, D., Verrier, E. (1996) Genetic variability of French dairy breeds estimated from pedigree information. *Interbull Bulletin*, **14**, 49-54.
- Moreno, A. (1998) Optimización de la respuesta a la selección en "Mus musculus" con consanguinidad restringida. PhD Thesis, Univ. Complutense, Madrid. Colección Tesis Doctorales, T22287.
- Pérez-Enciso, M. (1995) Use of the uncertain relationship matrix to compute effective population size. *J. Anim. Breed. Genet.*, **112**, 327-332.
- Valera, M., Molina, A., Gutiérrez, J.P., Gómez, J., Goyache, F. (2005) Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain. *Livest. Prod. Sci.*, **95**, 57-66.
- Wright, S. (1969) *Evolution and the Genetics of Populations: The Theory of Gene Frequencies*. Vol. 2. University of Chicago Press, Chicago, USA.

DISCUSIÓN GENERAL



5. DISCUSIÓN GENERAL

Esta discusión general se ha estructurado en dos bloques que, por su contenido, merecen ser discutidos de forma independiente. En el primero, se analiza la estructura genética de la población de caballo de Pura Raza Árabe Español y sus poblaciones derivadas. Este apartado se ha subdividido a su vez en otras dos partes, contemplando la estructura genética de la población inicial de forma independiente por un lado y la estructura de las razas derivadas por otro, dado que la segunda presenta además aportaciones metodológicas nuevas al estudio de poblaciones sometidas a migración. El segundo bloque de la discusión se refiere al desarrollo de nuevas metodologías de estimación del censo efectivo y es consecuencia de las deficiencias detectadas en los métodos disponibles al efecto en el momento de iniciar el primero. Se analizará el comportamiento del nuevo parámetro en poblaciones tipo, tanto simuladas como reales. Finalmente, se comentarán las ventajas de la nueva metodología así como los aspectos de la misma que quedan por resolver.

5.1. ESTRUCTURA GENÉTICA DEL CABALLO DE PURA RAZA ÁRABE ESPAÑOL Y SU INFLUENCIA EN RAZAS DERIVADAS

5.1.1. Estructura Genética del caballo de Pura Raza Árabe Español

El conocimiento de la variabilidad genética de una población constituye un paso previo necesario antes de tomar decisiones de selección o conservación sobre la misma. Entre otras opciones disponibles, la variabilidad genética de poblaciones que disponen de un registro genealógico puede ser estudiada mediante herramientas que analizan el flujo de los genes a través de las genealogías, proporcionando información relevante sobre la contribución de los fundadores y de consanguinidad. Estos parámetros son tanto más valiosos cuanto más profunda es la información de pedigrí de los individuos presentes en la población, por lo que el conocimiento de la cantidad y calidad de información disponible es en sí mismo importante.

5.1.1.1. Profundidad de pedigrí

La raza Árabe está distribuida mundialmente, existiendo diversas poblaciones con Libro Genealógico propio. Debido al intercambio que existe entre animales de raza Árabe de los distintos Libros Genealógicos, en muchas ocasiones las genealogías de los individuos que se importan no son completadas hasta su máxima generación conocida. Así, numerosos individuos que se registran en el Libro Genealógico del caballo Árabe Español, ven mermado su pedigrí, apareciendo como fundadores individuos que no los son en otras bases de datos extranjeras. Idealmente debería existir un Libro Genealógico único, donde se diesen de alta todos y cada uno de los individuos de raza Árabe, independientemente de su lugar de nacimiento. Así, podrían quedar registrados automáticamente todos los procesos de migración entre Libros Genealógicos Nacionales. Aunque existe una iniciativa por parte de la Arabian Horse Association (Asociación Mundial de Caballo Árabe) para recopilar la información procedente de todos los Libros

Genealógicos existentes, por dificultades en la gestión de los datos, la información reunida todavía no resulta apropiada para realizar análisis genealógicos. Por tanto, dado que no existe aún la posibilidad de analizar globalmente la población del caballo Árabe mundial, inicialmente, nos hemos centrado en la población del caballo Árabe Español, utilizando únicamente la información recogida en la base de datos proporcionada por su Libro Genealógico.

El intervalo generacional medio para la raza fue de 10,5 años. Con el objeto de que los resultados no se refiriesen al pedigrí completo sino que, fuesen representativos de la población más actual, se tomó como población de referencia los individuos nacidos durante la última generación.

Comparativamente y atendiendo al número de generaciones equivalentes, la población actual presenta una menor profundidad de pedigrí (7,9) que la principal raza equina autóctona de fomento español (caballo de Pura Raza Español) con 8,3 generaciones equivalentes (Valera *et al.*, 2005). Sin embargo, encontramos similitudes con la calidad el pedigrí del caballo Árabe Francés, en el que se conocen el 90% de los ascendientes en la cuarta generación, manteniéndose este mismo porcentaje para nuestra población hasta la sexta generación.

5.1.1.2. Relación entre las líneas paternas tradicionales y el análisis relativo a fundadores

El caballo Árabe Español tiene un origen diverso debido a la importación de caballos y yeguas de otros países. Con las primeras importaciones, realizadas por el llamado “Ministerio de Guerra” (actual Ministerio de Defensa) se trajeron principalmente caballos llamados del “Desierto”, procedentes de lo que podemos identificar geográficamente como Oriente Medio (principalmente Península Arábiga). También destacaron, desde el primer momento, las importaciones realizadas desde Polonia. En cuanto a la cría privada, fueron importantes las importaciones del Duque de Veragua con animales procedentes de Inglaterra, hecho que ha motivado que la procedencia de una de las líneas paternas tradicionales, el semental *Razada*, tenga su origen en una yegua llamada *Ranya* de Crabbet Park, un conocido núcleo de caballos Árabes en Inglaterra. Estas importaciones llevadas a cabo por el Duque de Veragua, junto con las adquisiciones del Marqués de Domecq, hizo que se creara una ganadería de PRÁ de alta calidad, cuya influencia se ha mantenido hasta la actualidad (Steen, 2006). Los sementales de las otras tres líneas paternas tienen su origen en el Oriente Medio (*Seanderich* y *Bagdad*) y en Polonia (*Ursus*).

La definición de estas líneas paternas tradicionales no tiene una justificación científica sino histórica, y, de alguna manera, también forma parte de la cultura general que rodea al mundo del caballo Árabe. Sin embargo, el análisis realizado agrupando la contribución de los fundadores por su procedencia, parece corroborar esta composición de líneas paternas tradicionales. Así, cabe destacar, la contribución de la población fundadora con origen en Oriente Medio, seguida de la población polaca, inglesa y egipcia. Actualmente se define también una contribución de casi un 10% de caballos fundadores nacidos en España. Esta

contribución sólo es superada por los caballos originarios de Oriente Medio, Polonia e Inglaterra.

Por otro lado, la variabilidad genética de esta población es también un reflejo de las circunstancias históricas que envolvieron a España en el siglo XX. La Guerra Civil, que produjo una disminución de los efectivos y el aislamiento de España por el régimen político, han hecho que los individuos que han aportado un 60% de la variabilidad presente en la población, sean los nacidos durante la época comprendida entre los años 1925 y 1944. Esto, unido a que las líneas paternas tradicionales del caballo Árabe Español se crearon en esas fechas, ha producido el uso abusivo de determinados ejemplares (tan solo 5 ejemplares serían necesarios para explicar el 50% de la variabilidad genética en la población actual). Podemos afirmar que La Guerra Civil en España provocó un cuello de botella en la población de Caballo Árabe, que se ha visto reflejado en el valor de los parámetros genealógicos que miden la variabilidad genética. Así, se ha encontrado un número efectivo de fundadores de 39,5 para la población de referencia, lo que supone un 4,6% del número total de fundadores encontrados, menos que en el caballo Árabe Francés (Moreaux *et al.*, 1996), aunque la comparación debe hacerse con cautela debido a las diferencias en la calidad del pedigrí. En cuanto al número efectivo de ancestros, que tiene en cuenta los cuellos de botella, fue de tan solo 13 para los animales nacidos en la última década. El número de genomas fundadores, calculado posteriormente a la redacción del capítulo correspondiente de esta memoria, fue aún inferior, con un valor de 6,6.



Ghandy, semental PRá, nacido en 1931, hijo de *Ursus*. Reproducción de 1950 (Yeguada Militar de Córdoba).



Nana-Sahib, Semental, nacido en 1934, hijo de *Razada*. Reproducción de 1950 (Yeguada Militar de Córdoba).

5.1.1.2.1. Estructuración de la población en cuanto a las líneas paternas estudiada mediante los índices de fijación de Wright

En cuanto a la diferenciación de la población por las diferentes líneas paternas descritas, el uso de determinados individuos considerados representativos de la línea (*Ursus*, *Seanderich*, *Razada* y *Bagdad*), pero no siempre descendientes directos de cada una de ellas, junto con la “sangre” materna, influenciada a su vez

por su “sangre” paterna y el cruzamiento constante entre las líneas descritas, conllevó la ausencia de diferenciación clara a nivel genealógico de las cuatro líneas estudiadas. Así, solo un 1% (F_{ST}) de la diferenciación genética de las 4 subpoblaciones identificadas se debe a que los individuos se asignen a una línea paterna u otra. Además, el uso generalizado de determinados sementales y las facilidades proporcionadas tradicionalmente por la Yeguada Militar con el servicio de sus sementales a través de las “Paradas de Sementales” (concentraciones públicas de sementales distribuidas por todo el territorio nacional, puestas a disposición del ganadero para cubrir las yeguas, previa solicitud, por un coste económico) pueden haber sido una de las causas más importantes de la falta de diferenciación genética en las teóricas subpoblaciones.

5.1.1.3. Las líneas de cría actuales y el análisis relacionado con la consanguinidad

La estructuración en líneas paternas también se refleja en las preferencias de cría que se pueden observar hoy día en España. De forma general podemos considerar internacionalmente cuatro tipos de “líneas de cría” importantes (AHA, 2008), de las que en España se dan principalmente tres:

- La línea Española (“Pure Spanish”) o animales de tipo Kuhaylan (fuerza). Son animales proporcionados, de cuello fuerte y arqueado y están bien desarrollados. Destacan por poseer buen carácter. Algunos ejemplos de esta línea son: los animales de Yeguada Militar (reseñar el semental *Garbo* y la yegua *Estopa*), los animales criados en la ganadería “La Aldara” (Sevilla) de Don Diego Méndez, la ganadería Ybarra y la de “Ses Planes” de Doña Marieta Salas (como ejemplo el semental *El Perfecto*).
- La línea Egipcia o animales de tipo Saqlaui (belleza). Son animales con cabeza y cuello estilizados, ojos pequeños y normalmente no destacan en movimiento. Suelen tener mal carácter.
- La línea Polaca. El caballo de esta línea presenta un cuello recto y destaca en movimientos. En Polonia la mayoría de los caballos Árabes se utilizan para carreras de hipódromo, lo que les permite probar su valor funcional en relación con la resistencia.
- La línea Rusa. Es de nueva creación originada por el cruce de las líneas Egipcia y Polaca. Aún está poco presente en nuestro país.

Estas cuatro “líneas” no son específicas ni de país ni de ganadería, de manera que podemos encontrar animales de las diferentes líneas en los diferentes países. Normalmente la cría se hace en pureza dentro de línea, aunque muchos ganaderos realizan cruces entre ellas. Uno muy habitual es llamado “Golden Cross”, que es el cruce de una yegua PRÁ española “Pure Spanish” (preferentemente procedente de la yegua *Estopa*) con un semental egipcio. Esta tendencia a reproducir dentro de líneas provoca una subdivisión incompleta pero permanente que llevará a valores de consanguinidad superiores a los que cabría

esperar en relación a los coeficientes de coascendencia, hecho que se debe revelar al comparar los valores de relación media con los de consanguinidad.

Así, con respecto al coeficiente de consanguinidad, en la población de referencia se encontró una media del 9,8%, valor muy por encima del encontrado en la población de caballos Árabe Francés (3,1%, Moreaux *et al.*, 2006) o Polaca (5,3%, Głazewska y Jezierski, 2004). Este incremento se produce principalmente después de los años 70, donde observamos con más claridad que el coeficiente de consanguinidad es superior a la mitad del coeficiente de relación media, lo que indica un apareamiento entre individuos consanguíneos, valores que deberían ser similares bajo apareamiento aleatorio. Este hecho se ve reforzado por el valor positivo que toma el F_{IS} en esta población (0,016 y 0,019 cuando se subdivide respectivamente por líneas paternas y por áreas geográficas). En cualquiera de los dos casos los valores de F_{ST} son demasiado bajos, por lo que parecen estar insuficientemente justificados por la estructuración analizada, de manera que existe probablemente alguna subdivisión mayor, probablemente debida a las diferentes líneas de cría establecidas en España referidas en el párrafo precedente.

Finalmente, el cruce entre parientes parece haber sido excesivo y podría haberse evitado. Así, en la Figura 2 e) del Artículo 4 podemos ver que en la última década, la consanguinidad ha descendido. Este hecho se puede explicar en parte por el establecimiento de controles en los apareamientos y al aumento en el uso de animales PRá extranjeros como parentales de la población española. La misma conclusión se deduce del descenso en los valores de los coeficientes de consanguinidad en la última década cuando sólo 5 generaciones son consideradas (Tabla 1 del Artículo 1).

5.1.2. Estructura Genética de las razas derivadas del caballo Árabe Español

El interés del PRá también se manifiesta por su participación en la formación de otras muchas razas equinas a lo largo de la historia. Por ello, ha sido interesante determinar su influencia en la formación de sus razas derivadas en España. El segundo objetivo de esta Tesis Doctoral abarca el estudio conjunto de las tres razas derivadas del caballo de PRá en España, determinando la influencia de la raza Árabe, que es el denominador común en la formación del caballo Anglo-Árabe (A-á, cruce de Pura Sangre Inglés -PSI- y Pura Raza Árabe), el caballo Hispano-Árabe (H-á, cruce de caballo de Pura Raza Español -PRE- y Pura Raza Árabe) y el Caballo de Deporte Español (CDE, población que contempla gran cantidad de cruces incluidos aquéllos en los que participa el PRá). Por otro lado, todas estas poblaciones están sometidas al fenómeno de la migración, por lo que se ha analizado el comportamiento de los diferentes parámetros genealógicos que definen la estructura de una población y se han realizado aportes metodológicos útiles en escenarios de migración.

Se ha pretendido deducir las consecuencias de ignorar la genealogía de los animales puros en las poblaciones cruzadas. Así, por un lado, el análisis se ha realizado mediante el uso de las "bases de datos exclusivas" del Libro

Genealógico de cada raza, en las que sólo aparecen dados de alta los animales de la propia raza, apareciendo los progenitores procedentes de las razas parentales como fundadores. Por otro lado, los resultados así obtenidos se han comparado con los calculados tras realizar una reconstrucción del pedigrí, buscando los ascendientes de razas parentales en el Libro Genealógico de procedencia (“bases de datos completadas”).

Para los tres casos se ha establecido una población de referencia utilizando los individuos registrados en los últimos 12 años para el caso del caballo Anglo-Árabe, últimos 10 años para el caballo Hispano-Árabe y últimos 11 años para el Caballo de Deporte Español, haciendo coincidir esos periodos con los intervalos generacionales medios para cada una de las razas.

5.1.2.1. Profundidad de pedigrí y consanguinidad

De nuevo, el factor limitante es la escasa profundidad de pedigrí registrada en las bases de datos exclusivas de estas razas, lo que se agrava al tratarse de poblaciones cuyo Libro Genealógico en el que puede inscribirse descendientes de razas parentales (PRÁ, PRE y PSI), además de estar abierto a animales de la misma raza procedentes de diferentes países, ignorando las relaciones de parentesco entre estos individuos que aparecen como “fundadores” en las razas derivadas (A-á, H-á y CDE).

Así, los valores medios del número equivalente de generaciones completas pasaron de ser 2,54, 1,35 y 1,35 en las bases de datos exclusivas de las razas derivadas, a 5,60, 7,71 y 3,35 para el caballo A-á, H-á y CDE respectivamente, en las bases de datos completadas con la información genealógica de los individuos de razas parentales.

Del mismo modo, se incrementó el coeficiente de consanguinidad al reconstruir la genealogía, demostrando la dependencia de este parámetro de la cantidad de pedigrí conocida. Los valores para los pedigríes reconstruidos en las tres poblaciones aumentaron desde un 30% en el caso del A-á hasta un 300% en el caso del H-á.

5.1.2.2. Influencia de la raza Árabe en sus poblaciones derivadas

La influencia de la raza Árabe en sus poblaciones derivadas se llevó a cabo desde dos puntos de vista:

- Desde el análisis de la población fundadora, sumando las contribuciones de los animales fundadores por raza (usando el coeficiente de relación media de ellos con el resto de la población). Los resultados obtenidos mostraron que para la raza A-á e H-á, la contribución de la raza Árabe fue siempre inferior a la otra raza parental (PSI o PRE) y tendió a ser mayor por la vía materna. En el caso del CDE, la contribución de fundadores PRÁ fue inferior al 10%. Este método refleja las preferencias de los criadores en la creación de animales cruzados pero no representa el parecido (en cuanto % de genes que comparten) entre las poblaciones actuales de las razas puras y las derivadas.

- Utilizando la matriz de coascendencias establecida entre las poblaciones de referencia de los 6 Libros Genealógicos, tanto de las razas derivadas (A-á, H-á y CDE) como de las razas parentales (PRá, PRE y PSI). Los resultados obtenidos muestran las consecuencias de los distintos procesos selectivos de las poblaciones en el parecido de los animales de las distintas poblaciones actuales. Se calculó el valor de la coascendencia media entre las diferentes poblaciones. Teniendo en cuenta que la coascendencia entre dos individuos es la mitad de su relación genética media, se obtuvo que, por término medio, un individuo de la raza A-á comparte el 4,6% de sus genes con un caballo de la población actual de PRá, mientras que un individuo H-á comparte el 6,4% con otro de la población actual de PRá. El porcentaje de genes compartidos con la otra raza parental (PSI y PRE) es en ambos casos inferior. Estos resultados son contradictorios a los obtenidos anteriormente y pueden deberse a que los individuos que se usan para reproducción en razas derivadas y raza parentales son distintos. Es decir, los criadores de A-á buscan en los caballos PSI unas cualidades que no buscan los criadores en pureza y lo mismo puede decirse para el H-á con el PRE. Sin embargo, no existe ese interés diferencial en el caso de los caballos PRá. Destaca también el hecho de que dos individuos de la misma raza derivada comparten por término medio menos genes entre sí, que con uno de la actual población del caballo Árabe Español.

5.1.2.2. Estimación del censo efectivo en las poblaciones derivadas del caballo Árabe

Dado que el coeficiente de consanguinidad depende de la profundidad de pedigrí, para tratar de describir la situación de una población, en cuanto a tasa de consanguinidad, es necesario el uso del censo efectivo. La estimación del censo efectivo, al igual que el resto de parámetros genealógicos, se ve afectado por la estructura de la población en estudio. En nuestro caso nos encontramos ante poblaciones sometidas a inmigración y generalmente con poca profundidad de pedigrí.

Se calcularon los tamaños efectivos usando la metodología de Gutiérrez *et al.* (2003) a partir de la información recogida en el Libro Genealógico de cada una de las razas en estudio (las bases de datos exclusivas), obteniendo valores que oscilaron entre el 16,2 para el caballo H-á y el 64,5 para el caballo A-á (pasando por el 47,0 estimado para el CDE). Sin embargo, por la propia estructura de la información recogida en el Libro Genealógico de cada raza, que no completa la información genealógica de los fundadores de razas parentales, desconocemos la relación que existe tanto entre estos individuos como con los animales de la población a la que contribuyen. Asimismo, se está ignorando la política de apareamientos llevada a cabo en las razas parentales de las que proceden. Esto se pone de manifiesto sobre todo en el caso del H-á, donde la genealogía es muy escasa para los individuos poco consanguíneos y abundante para individuos con elevada consanguinidad (solo 84 de los 1.783 individuos de la población de referencia eran consanguíneos pero el valor medio era muy alto, 13,7%). Esto se

traduce en un elevado coeficiente de regresión traducido en un pequeño censo efectivo, que no tiene por qué reflejar la situación real de la población.

Se utilizó la misma técnica con las bases de datos completadas, obteniendo valores de censo efectivo mayores para el caso del A-á (109,5) y el CDE (135,5), datos que, siendo un verdadero reflejo de los incrementos de endogamia en la población, no son comparables a los obtenidos en poblaciones puras con una política de apareamientos restringida a individuos de la misma raza. Así, los valores obtenidos en poblaciones con inmigración no pueden ser utilizados en comparación con los de poblaciones cerradas si pretenden utilizarse para contrastar los censos reales y la política de apareamientos. Por otro lado, para el caballo H-á, el valor que obtuvimos con la genealogía completada fue negativo, artefacto de la alta consanguinidad presente en la generación $t-1$ (parental en su mayoría, 6,38%) con respecto a la generación t (raza derivada, 1,91%).

5.1.2.2.1. El tamaño efectivo equivalente sin migración

Con el fin de comparar las poblaciones derivadas de otras similares, en términos de censos reales y política de apareamientos, fue necesario desarrollar otras metodologías de estimación del tamaño efectivo.

Nuestro objetivo era tratar de solventar el problema usando el modelo isla para estimar un tamaño efectivo equivalente. El modelo isla asume que una población sometida a una tasa constante de inmigración termina finalmente alcanzando un equilibrio entre la deriva y la creación de nueva variabilidad genética, de manera que la consanguinidad permanece constante. Dado que una población que no presenta incremento de consanguinidad se corresponde matemáticamente con un censo efectivo infinito, se desarrolla una expresión que relaciona el valor de la consanguinidad, constante, con la tasa de migración y el censo efectivo que tendría esa población si no estuviera afectada por la migración. El modelo isla asume que la población es una de las infinitas islas existentes por lo que cada migrante se asume como no relacionado con ningún otro.

El modelo isla es por tanto perfectamente aplicable a las bases de datos exclusivas de las poblaciones derivadas en las que aparecen sistemáticamente fundadores nuevos no relacionados con los que llegaron previamente, y puede ser empleado para estimar tamaños efectivos equivalentes sin migración. Se obtuvieron así los valores de censo efectivo equivalentes utilizando la tasa de migración en cada una de las bases de datos exclusivas de la raza y asegurándose de que no había incremento en la tasa de consanguinidad por año de nacimiento en la población de referencia (Falconer y Mackay, 1996). Los censos efectivos equivalentes estimados de este modo fueron 39,1, 56,3 y 114,1 para A-á, H-á y CDE respectivamente, que pueden ser comparables a los obtenidos en poblaciones puras y sin migración ($55,4 \pm 4.96$ para el PRE y $39,1 \pm 4.19$ para el PRÁ; Cervantes *et al.*, 2008). Los valores mostrados para las razas puras se corresponden con tamaños efectivos realizados obtenidos con la metodología desarrollada en esta memoria y muestran que la población de A-á sería comparable al PRÁ y el H-á al PRE en lo que se refiere al resultado de la combinación del censo real y la gestión de apareamientos.

5.2. EL TAMAÑO EFECTIVO REALIZADO. ESTIMACIÓN DEL CENSO EFECTIVO A PARTIR DEL INCREMENTO DE CONSANGUINIDAD INDIVIDUAL

Como ha sido ya referido en la revisión bibliográfica, el valor del coeficiente de consanguinidad es dependiente de la cantidad de información disponible por lo que no permite describir la situación de la población en relación a su pérdida de variabilidad genética. Para evitar esta dependencia de la cantidad de información genealógica disponible, se ha optado alternativamente por medir su incremento por generación (tasa de consanguinidad o de endogamia), lo que conduce de forma inversa a la definición del parámetro conocido como censo o tamaño efectivo. Este parámetro, debido a su relación directa con la tasa de consanguinidad, se considera como un parámetro imprescindible en el campo de la conservación y genética de poblaciones (FAO, 1998; Duchev *et al.*, 2006). De este modo, el tamaño efectivo ha sido adoptado como el criterio para clasificar las poblaciones atendiendo al grado de deterioro de su variabilidad genética.

El tamaño o censo efectivo es difícil de determinar en poblaciones de animales domésticos, dada la enorme variabilidad que existe en tamaño, estructura o gestión de apareamientos, por lo que su cálculo o estimación se ha convertido en un importante tema objeto de investigación. Como se ha descrito en la revisión bibliográfica, cuando se dispone de suficiente información genealógica, se puede abordar el cálculo del incremento de consanguinidad entre dos generaciones y, por relación inversa, obtener el censo efectivo (N_e). El incremento de consanguinidad en la población ideal es constante, independientemente del número de generaciones transcurridas y por tanto, el censo efectivo puede ser calculado de forma sencilla en esta población que sirve así de referencia. Aunque existen aproximaciones teóricas para el cálculo del censo efectivo cuando se incumple alguna de las condiciones de la población ideal, las poblaciones reales de animales domésticos presentan una extremada variedad de combinaciones de situaciones de solapamiento y estructuras que hacen inútil su aplicación.

5.2.1. El incremento de consanguinidad individual

Durante el desarrollo del objetivo 4 de esta Tesis Doctoral se ha implementado una nueva metodología de cálculo del censo efectivo basada en el incremento de consanguinidad individual. Este parámetro, el incremento de consanguinidad individual ΔF_i , representa una estandarización del coeficiente de consanguinidad con relación a la profundidad de pedigrí de cada individuo, que se corresponderá con la consanguinidad de un individuo en la primera generación si el incremento total se mantiene constante. De esta forma, al contrario que el coeficiente de consanguinidad, ΔF_i no depende del número de generaciones transcurridas en el pedigrí del animal analizado.

Aunque para el cálculo de ΔF_i es preciso disponer de una medida del número de generaciones de pedigrí que se conoce de cada individuo, Maignel *et al.* (1996) propusieron una medida del número equivalente de generaciones completas, permitiendo la estandarización de la consanguinidad. Ya que ΔF_i sólo depende del coeficiente de consanguinidad y del número de generaciones conocidas, podemos

ignorar las razones por las que nuestra población se aleja de la población ideal de referencia; todas éstas van a estar acumuladas en el parámetro sin necesidad de tenerlas en cuenta en los cálculos.

Este nuevo parámetro ha sido utilizado para medir la depresión consanguínea (González-Recio, *et al.*, 2007; Gómez *et al.*, 2008) y permite trabajar con un valor de consanguinidad que no depende de la profundidad de pedigrí, de manera que, por ser un parámetro novel su valor está aún por explotar. En el tema que nos ocupa veremos cómo puede ser utilizado para estimar el censo efectivo poblacional.

5.2.2. El tamaño efectivo realizado

La existencia de un incremento de consanguinidad diferente para cada individuo proporciona directamente la idea de variabilidad de este parámetro. De esta forma, un parámetro fijo, útil en la descripción de poblaciones, se transforma en una variable a la que se asocia una distribución, siendo preciso escoger un descriptor de esta distribución como estimador del parámetro fijo (Ej. una medida de centralidad de la misma como pueda ser la media aritmética). Dado que para cada individuo se dispone de un incremento de consanguinidad individual, podríamos disponer igualmente de un tamaño efectivo individual para cada animal. Por desgracia, los animales con coeficiente de consanguinidad nulo nos conducirían a un valor infinito de censo efectivo. En consecuencia, es preferible trabajar sobre la distribución de los incrementos de consanguinidad individuales y hacer después la transformación pertinente.

Así, a partir del inverso del doble del valor medio de los incrementos de consanguinidad individuales, de la población de referencia que se desea estudiar, se define en esta tesis el parámetro que hemos llamado “Tamaño o Censo Efectivo Realizado”. El adjetivo “realizado”, que se ha añadido al nombre del parámetro, pretende resaltar el aspecto descriptivo del mismo, dado que analiza lo que ha ocurrido en la población desde su fundación hasta la actualidad, acumulando todos los acontecimientos acaecidos durante su historia que puedan haber influido en los valores finales de consanguinidad. Sin embargo, el tamaño efectivo realizado no pretende predecir el comportamiento de la misma población en el futuro, siendo sólo útil como parámetro predictivo si la política de apareamientos no ha cambiado desde su fundación, ni se prevé que vaya a cambiar en la población.

El censo efectivo realizado, por estar basado en incrementos de consanguinidad individual, refleja todos los fenómenos ocurridos en cada individuo de la población, como el solapamiento generacional, la selección y la subdivisión que pueda existir. Es consecuencia de la política de apareamientos llevada a cabo, de los posibles cuellos de botella y del tamaño de la población base, todos ellos, factores que se reflejan en las genealogías.

Como fue comentado al comienzo de esta tesis, el tamaño efectivo puede ser explicado desde el punto de vista del cambio en las frecuencias alélicas o desde el punto de vista de los incrementos en los coeficientes de consanguinidad. Ambos

puntos de vista son equivalentes en la población ideal pero difieren en las poblaciones reales, si los apareamientos no son al azar, es decir, si el parentesco entre los individuos que se cruzan es distinto del parentesco medio de los animales disponibles para la reproducción. Éste es un hecho habitual en las poblaciones reales de animales domésticos. Conviene destacar que el tamaño efectivo realizado pretende estimar el censo efectivo, contemplado desde el punto de vista de la endogamia.

Así pues, una ventaja de esta metodología es que puede ser aplicada en cualquier escenario en el que se disponga de datos genealógicos, en beneficio sobre otras metodologías, comúnmente utilizadas, que no conducen a resultados satisfactorios en situaciones similares. Por ejemplo, las expresiones basadas en las diferencias en el número de descendientes asumen que los incrementos de coascendencia son iguales a los incrementos de endogamia y proporcionan valores del tamaño efectivo cuya explicación se refiere al punto de vista de las frecuencias alélicas. De este modo, ignoran cualquier tipo de apareamiento preferencial, hecho habitual en poblaciones total o parcialmente subdivididas, como es el caso de las poblaciones de animales domésticos.

Las técnicas basadas en regresión, consideradas como las técnicas de referencia para el cálculo del tamaño efectivo en poblaciones de animales domésticos con pedigrí, dependen mucho de los cambios en la política de apareamientos producidos durante el período contemplado en la población de referencia. De este modo, si tradicionalmente se han gestionados los cruces por técnicas de mínima coascendencia y en el período referenciado los cruces pasan a ser al azar, se producirá un brusco incremento de la consanguinidad en poco tiempo, lo que llevaría a una elevada pendiente en la regresión y a una fuerte subestimación del censo efectivo. El caso opuesto es aún más grave, ya que si se evitan los cruces entre parientes cercanos, tras haber descuidado los apareamientos en el pasado, la pendiente de la recta de regresión puede llegar a ser negativa en el período analizado, provocando estimaciones negativas del censo efectivo. El tamaño efectivo realizado, sin embargo, no puede ser nunca negativo ya que los coeficientes de consanguinidad nunca lo son.

Un valor añadido del parámetro es la posibilidad de disponer de una medida de incertidumbre, pues el incremento de consanguinidad individual es tratado como una variable, disponiéndose también de una varianza de la variable. Es posible así encontrar la varianza de la media de la variable teniendo en cuenta que el número de individuos independientes es precisamente el propio estimador del tamaño efectivo.

5.2.3. Ensayo del parámetro en poblaciones simuladas

La ventaja de la simulación es el conocimiento que se tiene del verdadero valor del parámetro que se desea estimar. Cuando se simula un apareamiento completamente aleatorio, en una población de censo constante estructurada en generaciones discretas, el valor estimado debe coincidir con el que se espera en la población ideal, con una leve corrección por el hecho de excluirse la autofecundación. Además de poder conocer el verdadero valor en esta situación,

disponemos también del verdadero valor cuando se incumple alguna de las condiciones de la población ideal. De este modo, la metodología fue puesta a prueba en circunstancias que recreaban situaciones habituales de las poblaciones de animales domésticos, como son la subdivisión, el solapamiento o cuando el número de descendientes por individuo no sigue la esperada distribución de Poisson.

El parámetro se comportó de forma excelente en todos los casos aunque conviene hacer algunas matizaciones:

- Existe una ligera sobreestimación del verdadero valor del parámetro que se va reduciendo a medida que avanzan las generaciones. Este retraso ha sido ya referido por Pérez-Enciso (1995) argumentando la existencia de dos sexos como causa y por tanto, imposibilitando la autofecundación. Parece entonces aconsejable restringir la población utilizada a aquellos individuos que dispongan de un número significativo de generaciones que podría situarse en torno a cinco.
- La variabilidad del parámetro es dependiente de su verdadero valor, presentando un importante efecto de escala. Por tanto, es de esperar que los estimadores tengan menos error cuando el censo efectivo de la población sea bajo.
- En el caso de la subdivisión se comprueba cómo el parámetro se ajusta de acuerdo a la historia completa de la población. De este modo, si se desea conocer el efecto de un cambio de política de apareamientos, conviene utilizar un pedigrí truncado en el momento en el que este cambio fue instaurado.

5.2.4. Ensayo del parámetro en poblaciones reales

La comprobación de la eficacia de esta metodología en poblaciones reales resulta difícil, ya que desconocemos el verdadero valor de censo efectivo. Para obtener una buena idea de la calidad del parámetro, se ha probado en un amplio abanico de situaciones. En primer lugar, se han elegido distintas bases de datos reales que representaban diversas políticas de apareamiento, como ejemplo de poblaciones reales subdivididas, apareamiento constante de mínimo parentesco, cambios en las políticas de apareamiento, poblaciones bajo selección artificial o sometidas a inmigración. Además, en cada una de ellas se ha calculado el parámetro en diferentes poblaciones de referencia definidas en función de períodos de nacimiento combinado con la profundidad de pedigrí. El resultado ha sido más que satisfactorio pudiéndose destacar algunos hechos:

- El parámetro muestra una estabilidad extraordinaria, presentando rangos muy estrechos de valores, independientemente de la subpoblación de referencia escogida dentro de cada población.
- Ofrece valores muy razonables cuando se han producido cambios en la política de apareamientos, que conducen a tendencias negativas del coeficiente de consanguinidad en el período analizado. Contrariamente los métodos basados en regresión proporcionan valores negativos. Este

hecho ha sido comprobado en el caballo de Pura Raza Español y en la Estirpe Cartujana del PRE.

- En ocasiones, en el período en el que se produce un descenso de la consanguinidad, se produce también paralelamente un descenso en el número de generaciones conocidas como consecuencia de una pequeña tasa de inmigración, como en el caballo de Pura Raza Árabe Español (ver Figura 2 e) del Artículo 4). Podemos observar que, en este caso, la regresión de la consanguinidad sobre el número de generaciones equivalentes devolverá un valor positivo pero por puro azar, por lo que su magnitud puede tomar cualquier valor. En el ejemplo señalado el valor obtenido por regresión fue de 18,1, frente al 39,1 obtenido con la nueva metodología, valor mucho más razonable y comparable con los obtenidos en otros períodos para la misma raza.
- Su comportamiento en escenarios de migración también es correcto. En estas situaciones se da un proceso constante de creación de variabilidad genética, por lo que el valor esperado del parámetro ha de ser muy alto. Así sucede en las estimaciones del tamaño efectivo realizado que se obtienen, en contraste con los valores obtenidos mediante técnicas basadas en regresión. Sin embargo, la migración conlleva que el número de generaciones conocidas sea muy bajo. El comportamiento del parámetro en esta situación no sería aceptable, como se desprende del ensayo llevado a cabo en poblaciones simuladas. Así lo refleja el elevadísimo error estándar que se obtiene en estas situaciones.

5.2.4.1. El tamaño efectivo en el caballo de Pura Raza Árabe Español

El nexo de unión entre los dos principales bloques de que consta este trabajo de Tesis Doctoral es la estimación del tamaño efectivo en el caballo de Pura Raza Árabe Español. El censo efectivo en la población se estimó con las diversas metodologías aplicadas durante la elaboración del quinto objetivo. Se utilizaron diversas poblaciones de referencia mediante la combinación de restricciones de pedigrí con los años de nacimiento de los individuos. Los valores obtenidos mediante regresión oscilaron entre 30,6, incluyendo la base de datos completa sin ninguna restricción, y los 11,4 utilizando como población de referencia individuos nacidos entre 1986 y 1997, con un número de generaciones equivalentes completas mayor o igual a 8. Los valores basados en la varianza del tamaño familiar presentaron cifras mucho mayores, oscilando entre 142,7, para toda la población, a 569,9 para individuos nacidos entre 1986 y 1997. Los reducidos valores obtenidos, mediante los métodos de regresión se deben a lo mencionado en el párrafo anterior, y pueden observarse en la Figura 2 e) del Artículo 4, hecho que puede explicarse por la reducción que se ha producido en los últimos años en la consanguinidad debido al mayor control en la gestión de los apareamientos y al uso de animales PRÁ de Libros Genealógicos extranjeros. Así, obtenemos un valor de tamaño efectivo de 18,1 si consideramos los animales nacidos en la última década. Por el contrario, el método basado en el tamaño familiar produce una sobreestimación del parámetro, dejando de lado la posible subdivisión presente en

la raza. Por ello, la nueva metodología desarrollada permite obtener un valor de censo efectivo mucho más razonable de 39,1 con un error estándar de 4,19. Este valor es comparativamente inferior al obtenido para el PRE (55,4), cuya población está cerrada aunque también tiene un censo 10 veces superior (MAPyA, 2003) y presenta un valor mayor que en poblaciones más pequeñas como la ganadería de Toro de Lidia de Don Juan Pedro Domecq y la Estirpe Cartujana del PRE. El valor en el PRÁ es reflejo de la situación actual en la población, en la que la gran afluencia de fundadores compensa la política de apareamientos y su situación histórica.

5.2.5. Limitaciones del tamaño efectivo realizado

Tal y como ha sido descrito, el ensayo llevado a cabo, tanto en poblaciones simuladas como reales, ha mostrado que el tamaño efectivo realizado es un excelente estimador del censo efectivo que se ha dado en una población y, por tanto, en nuestra opinión debería ser el método de elección cuando se trata de conocer el parámetro desde un punto de vista descriptivo, como por ejemplo, cuando se pretende clasificar una población en función de su grado de deterioro de variabilidad genética. Sin embargo, presenta limitaciones que conviene resaltar:

- Es un parámetro que de forma intrínseca recoge toda la historia de la población, desde que se dispone de registros genealógicos, por lo que determinadas actuaciones muy lejanas en el pasado podrían estar influenciando mucho su valor si la población o su gestión de apareamientos hubiese cambiado de forma significativa. Como ha sido ya comentado, una solución sería utilizar el pedigrí truncado desde la fecha posterior al cambio.
- Al basarse en informaciones genealógicas previas no puede ser utilizado para predecir el censo efectivo futuro si se sugiriese una modificación de la política de apareamientos, aunque podría simularse un pedigrí si las condiciones de la nueva política de apareamientos fueran muy concretas.
- Utiliza para su estimación los coeficientes de endogamia, por lo que no sirve para estimar el censo efectivo desde el punto de vista de los cambios en las frecuencias alélicas. Una posible solución a esta limitación podría explorarse mediante la extensión del método a los coeficientes de coascendencia en lugar de los coeficientes de endogamia.
- Otro problema radica en la necesidad de disponer de un número razonable de generaciones equivalentes conocidas, dado el retraso que parece producirse en la aparición de la consanguinidad en poblaciones dioicas sin autofecundación. Parece que es preciso un ajuste en este sentido.

El tamaño efectivo realizado como estimador del tamaño efectivo parece poseer propiedades muy interesantes, como se desprende de los análisis llevados a cabo en esta Tesis, tanto con poblaciones simuladas como reales. En nuestra opinión es hoy por hoy el mejor método disponible, pero su novedad deja abiertas muchas vías en las que se aconseja profundizar más en su investigación.

CONCLUSIONES



6. CONCLUSIONES

- I. La reducida variabilidad genética observada en el caballo de Pura Raza Árabe Español, ocasionada por el desequilibrio en la utilización de los reproductores y el cuello de botella acaecido durante la Guerra Civil parece estar compensándose en la actualidad, debido probablemente a un mayor control en la gestión de los apareamientos y al empleo de animales de Libros Genealógicos extranjeros como padres en las generaciones recientes, como se desprende del descenso en la consanguinidad media y en el número de generaciones equivalentes de los últimos años.
- II. La población de caballo Árabe Español presenta una subdivisión sólo parcialmente justificada por la existencia de líneas paternas reconocidas internacionalmente en esta población.
- III. A pesar de que nuestro caballo Árabe contribuye en menor proporción a las razas derivadas que otras razas puras, la población actual (animales nacidos en el último periodo correspondiente al intervalo generacional) de éstas razas derivadas comparte más genes con el el caballo de Pura Raza Árabe, dado que los sementales utilizados en la cría en razas derivadas y en la raza pura tienden a ser los mismos, al contrario que los sementales de las otras razas parentales.
- IV. El parámetro desarrollado aquí como tamaño efectivo equivalente sin migración permite detectar similitudes entre poblaciones cerradas (razas parentales puras) y abiertas (razas derivadas) desde el punto de vista de situación censal y la política de apareamientos. En este sentido el A-á sería comparable al PRá y el H-á al PRE.
- V. El tamaño efectivo realizado desarrollado en esta Tesis Doctoral se ha mostrado efectivo tanto en poblaciones simuladas como reales, presentando una gran estabilidad independientemente del número de generaciones consideradas y del período analizado. Además es el único estimador del censo efectivo que proporciona resultados coherentes en todo tipo de escenarios reales como la subdivisión, el solapamiento o los cambios fluctuantes en la política de apareamientos.
- VI. El tamaño efectivo realizado, aunque se comporta mejor en general que otros estimadores del censo efectivo, presenta pequeñas limitaciones que deben ser tenidas en cuenta en determinadas situaciones: sobreestima el verdadero valor con escasa profundidad de datos genealógicos, son necesarias modificaciones cuando sólo se quiere analizar un período concreto del total registrado y puede tener escaso valor predictivo.

REFERENCIAS BIBLIOGRÁFICAS



7. REFERENCIAS BIBLIOGRÁFICAS

- AHA. 2008. Arabian Horse Association. <http://www.arabianhorses.org>
- Ballou J.D.; Lacy, R.C. 1995. Identifying genetically important individuals for management of genetic variation in pedigreed populations. In: Ballou, J.D., Gilpin, M., Foose, T.J. (ed). Population management for survival and recovery: analytical methods and strategies in small population management. New York: Columbia University Press, pp: 76-111.
- Barreiros, R.; Barata, G. 1982. Efeito da consanguinidade em equinos. Rev. Port. Ciênc. Veter., 463, 167-177.
- Berthier, P.; Beaumont, M.A.; Cornuet, J.M.; Luikart, G. 2002. Likelihood-Based estimation of the effective population size using temporal changes in allele frequencies: A genealogical approach. Genetics, 160,741-751.
- Bowling, A.T.; Ruvinsky, A. 2000. The genetic resources and their conservation. In. "The Genetic of the horse". CAB International. UK, pp: 387-410.
- Boichard, D.; Maignel, L.; Verrier, E. 1997. The value of using probabilities of gene origin to measure genetic variability in a population. Genet. Sel. Evol., 29, 5-23.
- Caballero, A. 1994. Developments in the prediction of effective population size. Heredity, 73, 657-679.
- Caballero, A.; Toro, M.A. 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. Genet. Res., 75, 331-343.
- Caballero, A.; Toro, M.A. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conserv. Genet., 3, 289-299.
- Cavalli-Sforza, L.; Bodmer, W. 1971. The genetics of Human Populations. Ed. W.H. Freeman, San Francisco, Ed. CA, pp: 965.
- Cervantes, I.; Valera, M.; Gómez, M.D.; Medina, C.; Molina, A. 2006. Genetic parameters for the endurance raid in the Spanish Arab Horse. 57th European Association of Animal Production (Antalya, Turkey), pp: 351.
- Cervantes, I.; Goyache, F.; Molina, A.; Valera, M.; Gutiérrez, J.P. 2008. Application of individual increase in inbreeding to estimate realised effective sizes from real pedigrees. J. Anim. Breed. Genet., 125, 301-310.
- Chia, A.B.; Pollak, E. 1974. The inbreeding effective number and the effective number of alleles in a population that varies in size. Theor. Pop. Biol., 6, 149-172.
- Cunningham, E. 1991. Genética del Caballo Pura Sangre. Investigación y Ciencia, 178, 60-67.
- Crow, J.F.; Kimura, M. 1970. An Introduction to Population Genetics Theory. Harper and Row, New York.
- Duchev, Z.; Distl, O.; Groeneveld, E. 2006. Early warning system for loss of diversity in European livestock breeds. Archiv. Anim. Breed., 49, 521-531.

- Eding, H.; Meuwissen, T.H.E. 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *J. Anim. Breed. Genet.*, 118, 141-59.
- Falconer, D.S.; Mackay, F.C. 1996. *Introduction to Quantitative Genetics*. 4th edn, Longman Group Ltd. England.
- FAO. 1998. Secondary guidelines for the national farm animal genetic resources management plans: management of small populations at risk, FAO, Rome, Italy.
- Fenau, K. 1995. *El Pura Sangre Árabe*. Ediciones el Caballo, Barcelona. España.
- Fisher, R.A. 1930. *The genetical theory of natural selection*. Clarendon Press, Oxford. R.U.
- Gandini, G.; Bagnato, A.; Miglior, F.; Pagnacco, G. 1992. Inbreeding in the Italian Haflinger horse. *J. Anim. Breed. Genet.*, 109, 433-443.
- Głazewska, I.; Jezierski, T. 2004. Pedigree analysis of Polish Arabian Horses based on founder contributions. *Livest. Prod. Sci.*, 90, 293-298.
- Gómez, M.D.; Valera, M.; Cervantes, I.; Vinuesa, M.; Peña, F.; Molina, A. 2006. Development of a linear type trait system for Spanish Purebred horses (preliminary analysis). 57th European Association of Animal Production (Antalya, Turkey), pp 245.
- Gómez, M.D.; Valera, M.; Molina, A.; Gutiérrez, J.P.; Goyache, F. 2008. Assessment of inbreeding depression for body measurement in Spanish purebred (Andalusian) horses. *Livest. Sci.*, doi:10.1016/j.livsci.2008.08.007.
- González-Recio, O.; López de Maturana, E.; Gutiérrez, J.P. 2007. Inbreeding depression on female fertility and calving ease in Spanish dairy cattle. *J. Dairy Sci.*, 5744-5752.
- Gowe, R. S.; Robertson, A.; Latter, B. D. H. 1959. Environment and poultry breeding problems. 5. The design of poultry control strains. *Poultry Sci.*, 38, 462-471.
- Goyache, F.; Gutiérrez, J.P.; Fernández, I.; Gómez, J.; Álvarez, I.; Diez, J.; Royo, L.J. 2003. Using pedigree information to monitor genetic variability of endangered population: the Xalda sheep breed of Asturias as an example. *J. Anim. Breed. Genet.*, 120, 95-103.
- Gutiérrez, J.P.; Altarriba, J.; Díaz, C.; Quintanilla, R.; Cañón, J.; Piedrafita, J. 2003. Genetic analysis of eight Spanish beef cattle breeds. *Genet. Sel. Evol.*, 35, 43-64.
- Hill, W.G. 1979. A note on effective population size with overlapping generations. *Genetics*, 92, 317-322.
- Hill, W. G. 1981. Estimation of effective population size from data on linkage disequilibrium. *Genet. Res.*, 38, 209-216.
- James, J. 1972. Computation of genetic contributions from pedigrees. *Theor. Appl. Genet.*, 42, 272-273.
- Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge Univ. Press, Cambridge, Reino Unido.
- Klemetsdal, G.; Johnson, M. 1989. Effects of inbreeding on fertility in Norwegian Trotter. *Livest. Prod. Sci.*, 21, 263-272.

- Hamann, H.; Distl, O. 2008. Genetic variability in Hanoverian warmblood horses using pedigree analysis. *J. Anim. Sci.*, 86, 1503-1513.
- Lacy, R.C. 1989. Analysis of Founder Representation in Pedigrees: Founder Equivalent and Founder Genome Equivalents. *Zoo. Biol.*, 8, 111-123.
- Llamas, J.; Castelló, J.I.; Gil, A. 1992. *La Cría Caballar en España*. Ed. Darley SA (Barcelona, España).
- Lynch, M. 1996. A quantitative-genetic perspective on conservation issues. En "Conservation Genetics: Case Histories from Nature", ed. J.C. Avise & J.L. Hamrick. Chapman & Hall, New York.
- MacCluer, J.; Boyce, B.; Buke, L.; Weitzkamp, D.; Pfenning, A.; Parsons, C. 1983. Inbreeding and pedigree structure in Standardbred horses. *J. Heredity*, 74, 394-399.
- Maignel, L.; Boichard, D.; Verrier, E. 1996. Genetic variability of French dairy breeds estimated from pedigree information. *Interbull Bulletin*, 14, 49-54.
- Malécot, G. 1948. *Les Mathématiques de l'Hérédité*. Masson et Cie, Paris.
- MAPyA. 2003. Estudio y caracterización genética del sector equino en España. Ministerio de Agricultura Pesca y Alimentación. (<http://www.mapa.es/app/Equino/Informacion/Infsector.aspx?lng=e>). España.
- Maxwell, J. 1995. *Spanish Arabian Horse Families*. Alexander Heriot & Co. Ltd, England.
- Moureaux, S.; Verrier, E.; Ricard, A.; Meriaux, J.C. 1996. Genetic variability within French race and riding horse breeds from genealogical data and blood marker polymorphisms. *Genet. Sel. Evol.*, 28, 83-102.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York, pp: 512.
- Ollivier, L. 1981. *Éléments de génétique quantitative*. INRA et Masson, Paris, pp : 152.
- Oom, M. 1992. *O Cavalo Lusitano. Uma Raça em Recuperação*. Tesis Doctoral. Universidad de Lisboa.
- Pérez-Enciso, M. 1995. Use of the uncertain relationship matrix to compute effective population size. *J. Anim. Breed. Genet.*, 112, 327-332.
- Pirchner, F. 1985. Genetic structure of population. 1. Closed populations or matings among related individuals. In: Chapman, A. B. (ed.), *General and Quantitative Genetics*. World Animal Science A4 Elsevier, Amsterdam, The Netherlands.
- Quaas, R.L. 1976. Computing the diagonal elements of a large numerator relationship matrix. *Biometrics*, 32, 949-953.
- Radomska, M.; Burzynska, B.; Fiszdon-Stecko, K. 1984. Effect of inbreeding on the performance of Polish Arabian Horses. In 35th Annual Meeting of the EAAP. The Hague Netherlands. Paper No. 63, 10, pp: 2.
- Ron Garrido, L.; Birchmeier, A.N.; Munilla, S.; Cantet, R.J.C. 2008. Estimation of effective population size using bivariate discrete distributions for modelling family size in beef cattle. *Livest. Sci.*, 117, 43-51.

- Sölkner, J.; Filipcic, L.; Hampshire, N. 1998. Genetic variability of populations and similarity of subpopulations in Austrian cattle breeds determined by analysis of pedigrees. *Anim. Sci.*, 67, 249-256.
- Steen, A.K. 2006. A glorious task. A tale of the breed. Escandón Impresores. Seville. Spain.
- Valera, M.; Molina, A.; Satué, K. 1996. Influencia del nivel de consanguinidad en la población ganadera del caballo de P.S.á (Pura Sangre Árabe) en España. VI Congreso de Zootécnia. Nov. 1996. Evora.
- Valera, M.; Molina, A.; Gutiérrez, J.P.; Gómez, J.; Goyache, F. 2005. Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain. *Livest. Prod. Sci.*, 95, 57-66.
- Valera, M.; Cervantes, I.; Bartolomé, E. 2007. Otras razas equinas de Andalucía. En "Las Razas Ganaderas de Andalucía. Patrimonio Ganadero Andaluz". Volumen II, Junta de Andalucía. Sevilla, pp: 559-592.
- Van Eldik, P.; Van Der Waaij, E.H.; Ducro, B.; Kooper, A.W.; Stout, T.A.; Colenbrander, B. 2006. Possible negative effects of inbreeding on semen quality in Shetland pony stallions. *Theriogenology*, 65(6), 1159-1170.
- Vassallo, J.M.; Díaz, C.; García-Medina, J.R. 1986. A note on the population structure of the Avileña breed of cattle in Spain. *Livest. Prod. Sci.*, 15, 285-288.
- Wang, J. 2004. Application of the one-migrant-per-generation rule to conservation and management. *Conserv. Bio.*, 18, 332-343.
- Waples, R. S. 1989. A generalized approach for estimating effective population size 453 from temporal changes in allele frequency. *Genetics*, 121, 379-391.
- Weitkamp, L.; MacCluer, J.; Guttormsen, S.; McKnight, J.; Wert, N.; Witmer, J.; Boyce, P. and Egloff, J. 1982. Genetics of Standard-bred stallion reproductive performance. *J. Reprod. Fertil. (Suppl.)*, 32, 135-142.
- Wilkens, J.; Presinger, R.; Kalm, E. 1990. Effect of inbreeding on performance traits of riding horses based on competition results of the races "Holsteiner" and "Trakhener" warmblood. Institut für Tierzucht und Tierhaltung der Christian-Albrechts Universität zu Kiel Olshausentrabe, 40, D-2300. Kiel 1.
- Wright, S. 1922. Coefficient of inbreeding and relationship. *Am. Nat.*, 56, 330-338.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics*, 16, 97-159.
- Wright, S. 1938. Size of population and breeding structure in relation to evolution. *Science*, 87, 430-431.
- Wright, S. 1939. Statistical genetics in relation to evolution. *Exposés de Biométrie et de Statistique Biologique*. Herman & Cie, Paris, France.
- Wright, S. 1951. The genetical structure of populations. *Ann. Eugen. Lond.*, 15, 323-354.
- Wright S. 1969. Evolution and the Genetics of Populations: The Theory of Gene Frequencies, Vol. 2. University of Chicago Press, Chicago, USA.
- Wright, S. 1977. Evolution and the genetics of populations. Vol.3: Results and Evolutionary Deductions. The University of Chicago Press. Chicago.

- Wright, S. 1978. Evolution and the genetics of populations: Vol. 4. Variability within and among natural populations. University of Chicago Press: Chicago. USA
- Zechner, P.; Sölkner, J.; Bodo, I.; Druml, T.; Baumung, R.; Achmann, R.; Marti, E.; Habe, F.; Brem, G. 2002. Analysis Diversity and population structure in the Lipizzan horse bred based on pedigree information. Livest. Prod. Sci., 77, 137-146.

ANEXO



LISTADO DE PUBLICACIONES QUE HA DADO LUGAR LA PRESENTE TESIS DOCTORAL

PUBLICACIONES EN REVISTAS ISI

Título: "Population history and genetic variability in the Spanish Arab Horse assessed via pedigree analysis".

Autores: Cervantes, I.; Molina, A.; Goyache, F.; Gutiérrez, J.P.; Valera, M.

Revista: Livestock Science, 113: 24-33, 2008.

Título: "Individual increase in inbreeding allows estimating realised effective sizes from pedigrees".

Autores: Gutiérrez, J.P.; Cervantes, I.; Molina, A.; Valera, M.; Goyache, F.

Revista: Genetic, Selection and Evolution, 40: 359-378, 2008.

Título: "El cociente entre incrementos de endogamia y de coascendencia como medida de subdivisión poblacional. Resultados preliminares".

Autores: Cervantes, I.; Goyache, F.; Gutiérrez, J.P.

Revista: ITEA, 104(2) pp: 303-307, 2008.

Título: "Application of individual increase in inbreeding to estimate realised effective sizes from real pedigrees".

Autores: Cervantes, I.; Goyache, F.; Molina, A.; Valera, M.; Gutiérrez, J.P.

Revista: Journal of Animal Breeding and Genetics, 125: 301-310.

Título: "Genealogical analyses in open populations: the case of three Arab-derived Spanish horse breeds"

Autores: Cervantes, I.; Gutiérrez, J.P.; Molina, A.; Goyache, F.; Valera, M.

Revista: Journal of Animal Breeding and Genetics, submitted.

PUBLICACIONES DE CARÁCTER TÉCNICO

Título: "Primera valoración genética para la disciplina de raid en el caballo de Pura Raza Árabe Español".

Autores: Cervantes, I.; Molina, A.; Jódar, J.; Valera, M.

Revista: FEAGAS (Federación española de ganado selecto), 29:148-152, 2006.

LIBROS Y CAPÍTULOS DE LIBROS

Título: "Catálogo de Jóvenes Reproductores Recomendados 2006".

Autores: Escribano, C.J.; García, I.; Castellanos, M.; Valera, M.; Molina, A.; Roder, A.; Gómez, M.D.; Cervantes, I.; Bartolomé, E.

ISBN: 978-84-491-0771.

Editorial: Ministerio de Agricultura Pesca y Alimentación. Madrid. 2007.

Título: "Otras razas equinas. En: Patrimonio Ganadero Andaluz. Las razas ganaderas de Andalucía" Volumen II.

Autores: Valera, M.; Cervantes, I.; Bartolomé, E.

Editorial: Junta de Andalucía. 2007. Pp: 559-592.

Título: "Catálogo de Jóvenes Reproductores Recomendados 2007".

Autores: Valera, M.; Molina, A.; Roder, A.; Gómez, M.D.; Cervantes, I.; Bartolomé, E.; Escribano, C.J.; García, I.; Castellanos, M.

ISBN: 978-84-491-0819-8.

Editorial: Ministerio de Agricultura Pesca y Alimentación. Madrid. 2007.

CONGRESOS NACIONALES E INTERNACIONALES

Título: "Situación de las principales líneas genéticas tradicionales del Pura Raza Árabe en España".

Autores: Cervantes, I.; Sánchez, M.J.; Medina, C.; Membrillo, A.; Molina, A.; Valera, M.

Presentado en: IX Jornadas Científicas de Veterinaria Militar. 1-3 de Junio, Madrid, 2005.

Título: "Breeding evaluation of endurance horses in Spain (preliminary results)".

Autores: Cervantes, I.; Gómez, M.D.; Bartolomé, E.; Azor, P.J.; Molina, A.; Valera, M.

Presentado en: 9th Congress of the World Equine Veterinary Association. 22-26 January, Marrakech, 2006.

Título: "Genetic parameters for endurance ride in the Spanish Arab Horse".

Autores: Cervantes, I.; Valera, M.; Gómez, M.D.; Medina, C.; Molina, A.

Presentado en: 57th Annual Meeting of the European Association for Animal Production (EAAP). 17-20 September, Antalya, Turkey, 2006.

Publicado en el libro de resúmenes de 57th EAAP, pp: 351 (ISBN: 978-90-8686-003-6).

Título: "Demographic and reproductive parameters in the Spanish Arab horse".

Autores: Cervantes, I.; Gómez, M.D.; Molina, A.; Medina, C.; Valera, M.

Presentado en: 57th Annual Meeting of the European Association for Animal Production (EAAP). 17-20 September, Antalya, Turkey, 2006.

Publicado en el libro de resúmenes de 57th EAAP, pp: 246 (ISBN: 978-90-8686-003-6).

Título: "Modelos para la valoración genética de la disciplina de concurso completo de equitación a partir del control de rendimientos en las pruebas de selección de caballos jóvenes".

Autores: González, M.A.; Cervantes, I.; Gómez, M.D.; Bartolomé, E.; Serradilla, J.M.; Valera, M.

Revista: ITEA (Información técnica económica agraria), Vol. Extra 28, Tomo II: 438-440, 2007.

Título: "Accounting for migration rates to compute effective population size in three Arab derived Spanish horse breeds".

Autores: Cervantes, I.; Gutiérrez, J.P.; Goyache, F.; Bartolomé, E.; Molina, A.; Valera, M.

Presentado en: 58th Annual Meeting of the European Association for Animal Production (EAAP). 26-29 August, Dublin, Ireland, 2007.

Publicado en libro de resúmenes de 58th EAAP, pp: 154 (ISBN: 978-90-8686-045-6).

Título: "Morphological data analysis of Spanish Arab Horse aimed to define a line type trait system".

Autores: Cervantes, I.; Gómez, M.D.; Bartolomé, E.; Gutiérrez, J.P.; Molina, A.; Valera, M.

Presentado en: 58th Annual Meeting of the European Association for Animal Production (EAAP). 26-29 August, Dublin, Ireland, 2007.

Publicado en libro de resúmenes de 58th EAAP, pp: 158 (ISBN: 978-90-8686-045-6).

Título: "Genetic parameters for young eventing competition in Spain: correlation between dressage, jumping, cross and conformation".

Autores: Cervantes, I.; Bartolomé, E.; Gómez, M.D.; Medina, C.; González, M.A.; Valera, M.

Presentado en: 58th Annual Meeting of the European Association for Animal Production (EAAP). 26-29 August, Dublin, Ireland, 2007.

Publicado en libro de resúmenes de 58th EAAP, pp: 157 (ISBN: 978-90-8686-045-6).

Premio al mejor póster de la sección equina.

Título: "Accounting for the breed effect in the estimation of genetic parameters of horse eventing competition traits".

Autores: Cervantes, I.; Gutiérrez, J.P.; Molina, A.; Bartolomé, E.; Gómez, M.D.; Goyache, F.; Valera, M.

Presentado en: 59th Annual Meeting of the European Association for Animal Production (EAAP). 24-27 August, Vilnius, Lithuania, 2008.

Publicado en libro de resúmenes de 58th EAAP, pp: 287 (ISBN: 978-90-8686-074-6).

Título: "Subdivision of Spanish horse populations assessed by comparing effective sizes computed from individual increases in both inbreeding and coancestry"

Autores: Cervantes, I.; Valera, M.; Goyache, F.; Molina, A.; Gutiérrez, J.P.

Presentado en: 59th Annual Meeting of the European Association for Animal Production (EAAP). 24-27 August, Vilnius, Lithuania, 2008.

Publicado en libro de resúmenes de 58th EAAP, pp: 288 (ISBN: 978-90-8686-074-6).

Esta Tesis Doctoral ha sido parcialmente subvencionada por una beca de la Consejería de Educación y Ciencia de la Junta de Andalucía (BOJA , nº 120, 21/06/2004) y a través del contrato establecido entre la Asociación Nacional de Criadores de Caballos Árabes y la Universidad de Sevilla, que a su vez es financiado por la Dirección General de Recursos Agrícolas y Ganaderos del Ministerio de Medio Ambiente, Medio Rural y Marino.

